

VISIT...

LANZAROTE
Caliente.COM

**ИНСТИТУТ ФИЗИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ
РОССИЙСКАЯ АКАДЕМИЯ НАУК**

На правах рукописи

Григорьев Вениамин Юрьевич

**КОЛИЧЕСТВЕННЫЕ МОДЕЛИ «СТРУКТУРА – СВОЙСТВО»
ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ**

02.00.03 – органическая химия

02.00.04 – физическая химия

Диссертация

на соискание ученой степени доктора химических наук

Черноголовка 2013

Условные обозначения и сокращения

α – молекулярная поляризуемость

ΔG – свободная энергия Гиббса образования комплекса с водородной связью

ΔH – энтальпия образования комплекса с водородной связью

ACC (accuracy) – точность, характеристика бинарной классификации

AChE – ацетилхолинэстераза

AD – область применимости

ADME – адсорбция, распределение, метаболизм, выделение

AFP – адаптивное нечеткое распределение

ANN – искусственная нейронная сеть

C_a – свободноэнергетический протоноакцепторный дескриптор водородной связи

C_c – косинусный коэффициент

C_d – свободноэнергетический протонодонорный дескриптор водородной связи

cov - ковариация

C_v – коэффициент вариации

d – плотность вещества

D – фрактальная размерность

DA – дискриминантный анализ

DT – дерево решений

E_a – энтальпийный протоноакцепторный дескриптор водородной связи

E_c – Эвклидов коэффициент

EC_{50} – среднеэффективная концентрация вещества

E_d – энтальпийный протонодонорный дескриптор водородной связи

E_{HOMO} – энергия высшей занятой молекулярной орбитали

E_{LUMO} – энергия низшей свободной молекулярной орбитали

E_s – стерический параметр

F – критерий Фишера

FATS – синдром острой токсичности у рыб

GA – генетический алгоритм

GER - программирование экспрессии генов

HCA – иерархический кластерный анализ

I - степень ингибирования биосинтеза гиббереллина

IC₅₀ – концентрация вещества, вызывающая 50% ингибирование

ILP – индуктивное логическое программирование

K_{aw} – коэффициент распределения вещества в системе воздух – вода

KNN – метод k-ого ближайшего соседа

K_{ow} – коэффициент распределения вещества в системе н-октанол – вода

LC₅₀ – среднесмертельная концентрация вещества

LD₅₀ – среднесмертельная доза вещества

LDA – линейный дискриминантный анализ

LNO – кросс-валидация с выбором по N

logBB – десятичный логарифм отношения концентраций вещества в мозге и крови при пассивной диффузии через ГЭБ

LOO – кросс-валидация с выбором по одному

LR – логистическая регрессия

MCC – коэффициент корреляции Мэтьюза

MlogP – десятичный логарифм расчетного коэффициента распределения вещества в системе н-октанол – вода

MLR – множественный регрессионный анализ

MOA (mode of action) – тип токсического действия

n – число наблюдений

N_{Hacc} – число H-акцепторов

N_{Hdon} – число H-доноров

OECD – организация экономического сотрудничества и развития

P – давление насыщенного пара

PCA – анализ главных компонент

pK_a – отрицательный десятичный логарифм константы диссоциации кислоты

PLS – проекции на скрытые структуры

q – коэффициент линейной корреляции при кросс-валидации с выбором по одному

QSAR – количественная связь «структура – активность»

QSPR – количественная связь «структура – свойство»

r – коэффициент линейной корреляции

RF – метод случайного леса

s – стандартное отклонение

s_{cv} – стандартное отклонение при кросс-валидации с выбором по одному

SN (sensitivity) – чувствительность, характеристика бинарной классификации

SP (specificity)- специфичность, характеристика бинарной классификации

SS – метод структурного сходства

SVILP - индуктивное программирование на опорных векторах

SVM – метод опорных векторов

T – температура

t – время

T_c – коэффициент Танимото

АЛБТ – анализ локальной базовой токсичности

AMP – арифметическое среднее свойство

БД – база данных

ГМФТА – гексаметилфосфортриамид

ГЭБ – гематоэнцефалический барьер

ЛДРП – линейный дискриминантно-регрессионный подход

ЛРМПК - локальные регрессионные модели в перекрывающихся кластерах

ФАВ – физиологически активные вещества

Оглавление

Условные обозначения и сокращения	2
Введение	7
Глава 1. Современное состояние в области создания количественных моделей «структура – свойство» и «структура – активность» (на примере острой токсичности органических соединений)	13
1.1. Общая схема QSPR/QSAR моделирования	13
1.2. Объекты исследования и источники данных	15
1.3. Дескрипторы	19
1.4. Кластеризация молекул. Предобработка данных	22
1.5. Отбор дескрипторов	26
1.6. Статистические методы моделирования	28
1.7. Валидация моделей	33
1.8. Модели	39
1.9. Экспертные системы	73
Выводы по главе	76
Глава 2. Экспериментальная часть	78
2.1. Физико-химические данные	78
2.2. Биологические данные	78
2.3. Расчет, преобразование и отбор дескрипторов	79
2.4. Меры сходства химических соединений	81
2.5. Статистические методы исследования	82
2.6. Валидация QSPR/QSAR моделей	84
2.7. Методика проведения спектральных и калориметрических измерений	87
2.8. Планирование эксперимента и оценка термодинамических параметров водородной связи по данным экспериментальных измерений	88
2.9. Исходные вещества, реактивы и их квалификация	90
Глава 3. Количественные модели «структура – физико-химические свойства» органических соединений	91
3.1. Аддитивно-мультипликативная модель энтальпии и свободной энергии	

водородной связи. Монофункциональные соединения	91
3.2. Аддитивно-мультипликативная модель энтальпии и свободной энергии водородной связи. Полифункциональные соединения	116
3.3. Модель липофильности органических соединений	122
3.4. Компьютерный программный комплекс НУВОТ	134
Выводы по главе	140
Глава 4. Количественные локальные модели «структура – биологическая активность» органических соединений	142
4.1. Классификационные и регрессионные модели неполярного наркоза ...	142
4.2. Классификационные и регрессионные модели полярного наркоза	157
4.3. Линейный дискриминантно - регрессионный подход	182
4.4. Ретардантная активность солей четвертичного аммония	191
Выводы по главе	195
Глава 5. Количественные глобальные модели «структура – биологическая активность» органических соединений	197
5.1. Модель АЛБТ	197
5.2. Модель АМР	213
5.3. Модель ЛРМПК	225
5.4. Проницаемость ФАВ через гематоэнцефалический барьер	233
Выводы по главе	238
Глава 6. Перспективные разработки в области создания количественных моделей «структура – свойство» и «структура – активность»	240
6.1. Рекуррентная модель	240
6.2. Фрактальный дескриптор	251
6.3. Линейная динамическая модель	265
Выводы по главе	277
Заключение	278
Список литературы	280

Введение

Одной из фундаментальных задач химии является установление связи между структурой вещества и проявляемыми им свойствами. К настоящему времени в мире синтезировано несколько десятков миллионов химических соединений, большинство из которых относится к органическим веществам. Эти соединения могут быть охарактеризованы большим количеством свойств, общее число которых, по некоторым оценкам, приближается к 10000.

В сферу интересов человека вовлечено около 60 тысяч органических соединений, которые широко применяются в быту, медицине, на производстве и в сельском хозяйстве. Этот список ежегодно увеличивается примерно на 1000 наименований [1]. Для многих соединений отсутствуют данные об их физико-химических и биологических свойствах, например, о способности к образованию водородной связи или об острой токсичности. Но распределение Н-донорных и Н-акцепторных центров и величины их активности нужны при конструировании новых биологически активных соединений, а знание количественных характеристик токсичности необходимо для оценки опасности промежуточных и конечных веществ при планировании и проведении многостадийного органического синтеза.

Для экспериментального определения числовых характеристик свойств соединений необходимы большие материальные и временные затраты. Однако полученные к настоящему времени экспериментальные данные могут служить базой для разработки и дальнейшего использования математических моделей тех или иных свойств, созданных на основе, например, междисциплинарного подхода «количественная связь структура – свойство» (QSPR)/«количественная связь структура – активность» (QSAR). Компьютерное конструирование таких моделей требует гораздо меньше времени и материальных ресурсов, чем экспериментальные исследования, а созданные модели позволяют оценивать свойства неисследованных химических соединений и предсказывать новые структуры с заданными свойствами.

Работа выполнена в соответствии с планами научно-исследовательских работ Института физиологически активных веществ РАН «Компьютерный молекулярный дизайн, включая количественную связь структура-свойство (активность) и молекулярное моделирование взаимодействия лиганд-мишень для структурной генерации соединений с заданными свойствами» (№ госрегистрации 0120.0 504023) при финансовой поддержке Российского фонда фундаментальных исследований: проект № 97-03-33068-а «Молекулярное моделирование связи структура-активность с учетом трехмерности структур и их способности к образованию водородных связей» и Международного научно-технического центра: проект № 888 «Компьютерное предсказание химической экотоксичности» и проект № 3777 «Развитие и применение КССА моделей токсичности химических соединений по отношению к грызунам».

Цель исследования

Разработка количественных моделей «структура – свойство» и «структура – активность» с использованием дескрипторов водородной связи и других молекулярных дескрипторов для компьютерного дизайна органических соединений, включая прогнозирование физико-химических свойств и биологической активности неисследованных соединений и предсказание новых структур с заданными свойствами.

Задачи исследования

1. Разработка модели для количественной оценки энтальпии и свободной энергии Н-комплексов, создание единой энергетической шкалы Н-связи и расчет дескрипторов водородной связи.
2. Создание базы данных по термодинамическим характеристикам Н-комплексов и базы данных по Н-донорным и Н-акцепторным дескрипторам.
3. Разработка модели липофильности органических соединений.
4. Конструирование локальных и глобальных моделей «структура – активность» органических соединений.

5. Разработка рекуррентной модели, фрактального дескриптора и линейной динамической модели для установления количественной связи «структура – свойство» и «структура – активность».

Научная новизна

На основе литературных и собственных экспериментальных данных разработана физико-химическая аддитивно-мультипликативная модель энтальпии и свободной энергии Н-комплексов, сконструирована единая энергетическая шкала водородной связи и создана компьютерная программа для молекулярного дизайна органических соединений с заданной величиной Н-донорной и Н-акцепторной активности.

Создана модель липофильности органических соединений, основанная на линейной комбинации двух физико-химических параметров: молекулярной поляризуемости и Н-акцепторного дескриптора водородной связи.

С использованием молекулярных дескрипторов, в том числе дескрипторов Н-связи, сконструированы оригинальные кластерные, классификационные и регрессионные модели «структура – активность», предназначенные для компьютерного дизайна органических соединений, включая оценку острой токсичности, ретардантной активности и способности проникновения через гематоэнцефалический барьер неисследованных новых соединений и генерацию ряда структур с заданным уровнем биологической активности.

Впервые разработана рекуррентная модель для описания связи «структура – активность» в гомологических рядах органических соединений, отличающаяся от известных моделей отсутствием дескрипторов.

Впервые предложен, разработан и применен новый 3D дескриптор для оценки фрактальных свойств молекулярных гистограмм межатомных расстояний.

Впервые предложена, разработана и применена линейная динамическая модель для компьютерного молекулярного дизайна 3D структур с заданными свойствами.

Научно-практическая значимость

1. База данных по термодинамическим характеристикам образования Н-комплексов, насчитывающая свыше 13000 записей, может быть использована в качестве источника информации при проведении теоретических и практических исследований в области органической и физической химии.
2. База данных по протонодонорным и протоноакцепторным дескрипторам Н-связи, содержащая свыше 50000 записей, и компьютерная программа HYBOT, созданная на ее основе, могут служить для количественной оценки способности органических соединений к образованию водородной связи.
3. На основе сконструированных QSPR/QSAR моделей может проводиться оценка физико-химических свойств и биологической активности новых неизученных органических соединений.

Основные положения, выносимые на защиту

1. Аддитивно-мультипликативная модель энтальпии и свободной энергии Гиббса Н-комплексов позволяет количественно оценить протонодонорную и протоноакцепторную способность органических соединений.
2. QSPR модель липофильности дает возможность значительно расширить круг органических соединений, для которых могут быть рассчитаны дескрипторы Н-связи.
3. Модели неполярного и полярного наркоза на базе молекулярной поляризуемости и дескрипторов Н-связи, позволяющие оценить величины токсичности и вклады неспецифических и специфических взаимодействий.
4. Алгоритмы кластеризации/классификации на основе регрессионных моделей наркоза, предоставляющие возможность провести

кластеризацию/классификацию соединений без использования априорной информации о типе токсического действия каждого соединения.

5. Рекуррентная модель острой токсичности гомологических рядов органических соединений позволяет проводить корректную интерполяцию и экстраполяцию значений токсичности, а также оценивать токсикологическую близость живых организмов или химических соединений.
6. Глобальные модели «структура – активность», основанные на молекулярном сходстве, и физико-химических дескрипторах, дают возможность оценивать биологическую активность органических соединений с точностью, соответствующей точности исходных данных.
7. Фрактальный дескриптор гистограмм межатомных расстояний количественно оценивает сложность молекулярных структур в их геометрическом аспекте.
8. Линейная динамическая модель связи структуры и свойства, выраженных в виде двумерных кривых, позволяет не только рассчитать свойства по структуре, но также получить информацию о трехмерной структуре соединений с заданными свойствами.

Личный вклад автора

Все результаты диссертации получены лично автором или в соавторстве при его непосредственном участии. В работах, выполненных в соавторстве, вклад автора состоял в непосредственном участии во всех этапах работы, начиная от постановки задачи и завершая обсуждением результатов. Синтез, очистку и идентификацию новых физиологически активных соединений выполнили сотрудники ИФАВ РАН: Иванов А.Н., Соколов В.Б., Чепакowa Л.А., Брель В.К., Додонов М.В., Прошин А.Н., Чистяков В.Г. Биологические испытания на ретардантную активность провел Кокурин А.В. (ВНИИСБ РАСХН). Вклад автора в программный комплекс HUBOT, разработанный совместно с Раевским О.А., Трепалиным С.В. и Раевской О.Е. (ИФАВ РАН),

заклучался в создании термодинамической базы данных и базы дескрипторов Н-связи.

Автор выражает глубокую признательность доктору химических наук, профессору Раевскому О.А. за ценные научные консультации во время выполнения работы.

Апробация работы

Результаты исследования и основные положения работы были представлены на IX Всесоюзной конференции «Химическая информатика» (Черноголовка, 1992); на 12-ом Международном симпозиуме по медицинской химии (Базель, Швейцария, 1992); на 9-ом Европейском симпозиуме по количественным соотношениям структура-активность (Страсбург, Франция, 1992); на IV Российском конгрессе «Человек и лекарство» (Москва, 1997); на VI Российском конгрессе «Человек и лекарство» (Москва, 1999); на 4-ом Международном симпозиуме по компьютерным методам в токсикологии и фармакологии (Москва, 2007); на XVII Российском конгрессе «Человек и лекарство» (Москва, 2010); на 18-ом Европейском симпозиуме по количественным соотношениям структура-активность (Родес, Греция, 2010); на XVIII Российском конгрессе «Человек и лекарство» (Москва, 2011); на XIX Российском конгрессе «Человек и лекарство» (Москва, 2012).

Публикации: по теме диссертации опубликовано 40 печатных работ, включая 20 статей в рецензируемых российских и международных журналах, получено 3 свидетельства об официальной регистрации программ для ЭВМ.

Объем и структура диссертации Диссертация состоит из введения, обзора литературы, экспериментальной части, 4-х глав, посвященных результатам исследования и их обсуждению, заключения, а также списка литературы из 427 наименований. Работа изложена на 324 страницах машинописного текста, содержит 37 таблиц и 33 рисунка.

Глава 1. Современное состояние в области создания количественных моделей «структура – свойство» и «структура – активность» (на примере острой токсичности органических соединений)

1.1. Общая схема QSPR/QSAR моделирования

За последнее десятилетие опубликован ряд обзоров, освещающих те или иные аспекты применения QSAR при исследовании острой токсичности органических соединений. История использования, основные концепции и цели QSAR в токсикологии представлены в работе [2]. Современное состояние и перспективы освещены в публикации [3]. В обзорных работах [4-6] представлены модели количественной связи структура – токсичность для водных организмов, а в [5, 7-9] модели для земных организмов, в том числе, для млекопитающих. Вопросы, связанные с острой токсичностью лекарственных средств и их компьютерные модели, рассмотрены в публикации [10]. В обзоре [11] исследованы различные стороны количественного моделирования, включая 2D и 3D QSAR, выбор дескрипторов, валидацию моделей. Результаты расчетов острой токсичности органических соединений с применением искусственных нейронных сетей представлены в работах [12, 13]. Роль типов токсического действия при использовании QSAR проанализирована в публикации [14]. В обзоре [15] рассмотрены регрессионные модели токсичности различных соединений, действующих по радикальному механизму. В работе [16] представлены модели QSAR для водных организмов.

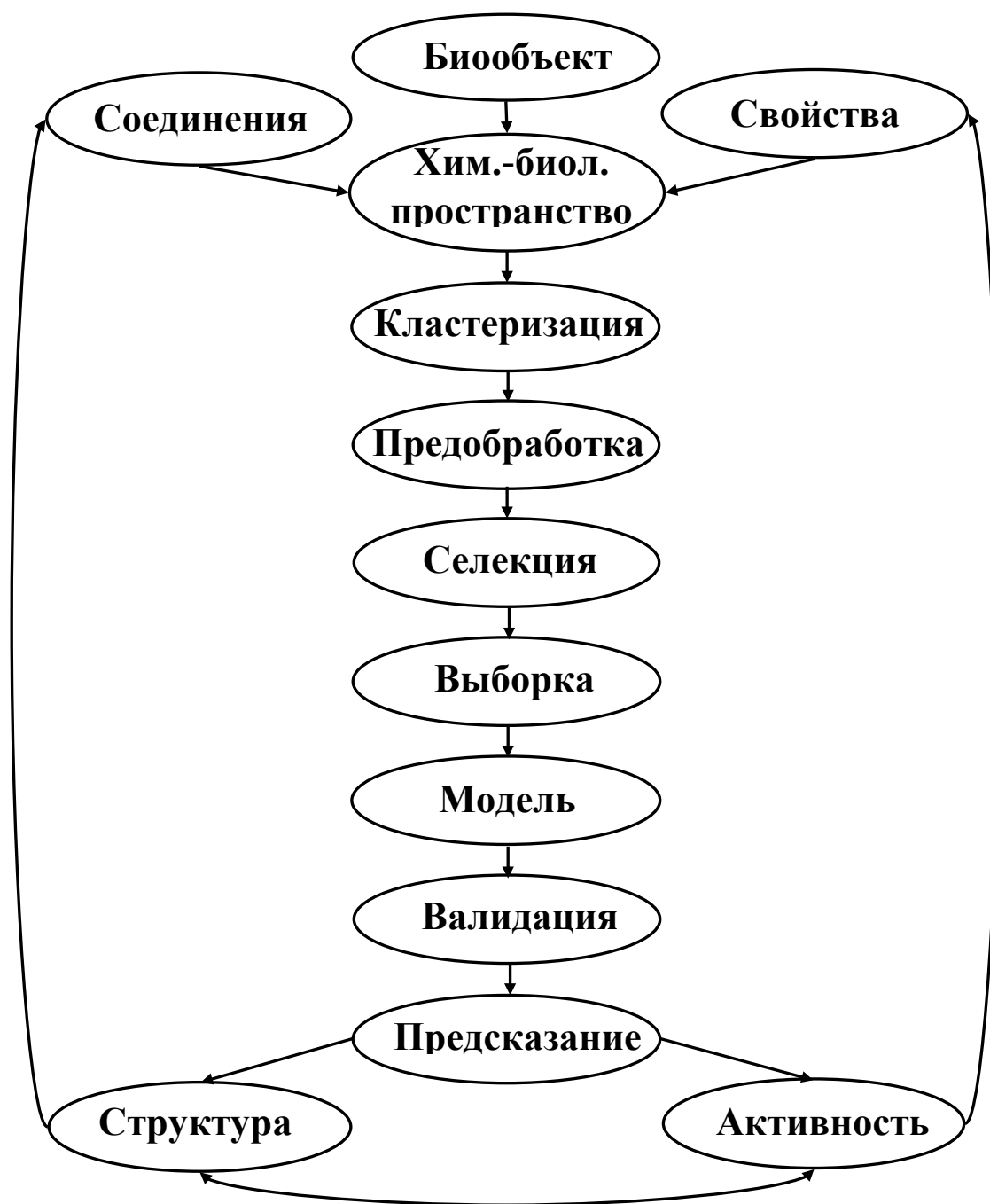


Рис. 1. Общая схема QSPR/QSAR моделирования

Обобщенная схема QSPR/QSAR моделирования, представленная на рис. 1, свидетельствует о сложном и многоэтапном пути построения QSPR/QSAR моделей. Ведущими «участниками» рассматриваемой схемы являются химические соединения, биологические объекты и связанные с ними свойства и токсичности. Их совокупность можно представить в виде химико-биологического пространства, геометрической интерпретацией которого являются точки (молекулы) в многомерном пространстве (дескрипторы, токсичности, биообъекты). Следующим необходимым шагом является кластеризация химического пространства с целью формирования выборки молекул и соответствующей группы зависимых (токсичности) и независимых (дескрипторы) переменных. Предобработка данных состоит в анализе пропущенных значений и проведении преобразования переменных, например, логарифмирование значений токсичности и автошкалирование дескрипторов. На этапе селекции решается ряд задач, включая снижение размерности пространства, отбор информативных дескрипторов и отбрасывание линейно зависимых переменных. В результате проведенных операций формируется обучающая выборка. На ее основе проводится моделирование зависимости токсичности от дескрипторов с использованием различных статистических подходов. Полученные модели подвергаются валидации. Она включает в себя анализ выбросов, оценку описательной и предсказательной способности зависимостей, а также определение их области применимости. С использованием полученных моделей может быть оценена токсичность новых соединений (прямая задача) или предсказаны структуры с заданными значениями токсичности (обратная задача).

1.2. Объекты исследования и источники данных

Основными биологическими объектами при изучении острой токсичности органических соединений являются водные и земные организмы. Среди водных организмов наиболее часто встречаются следующие типы:

инфузории *Tetrahymena pyriformis* [17-80, 214, 221, 239, 311], рыбы *Pimephales promelas* [17, 19-21, 52, 54-56, 79, 81-111, 182, 196, 197, 238], *Poecilia reticulata* [16, 22, 79, 80, 86, 108, 109, 112-115, 197], *Oncorhynchus mykiss* [86, 107, 109, 110, 116-124, 182, 229] и *Cyprinus carpio* [79, 80, 109, 182, 185, 186, 188, 190], бактерии *Vibrio fischeri* [17, 79, 80, 111, 125-142, 188, 218, 227, 310], членистоногие *Daphnia magna* [53, 54, 79-81, 107, 111, 122, 135, 143-155, 182, 186-190, 192, 193, 196, 198, 218, 229] и водоросли *Chlorella vulgaris* [17, 18, 80, 156-161, 195]. Наиболее часто в качестве объектов исследования среди земных организмов выступают грызуны: мыши [7, 162-173, 195, 227, 237, 241] и крысы [7, 123, 167-181, 183, 227, 240, 241].

Отдельные публикации посвящены изучению таких организмов, как земноводные *Bufo vulgaris formosus* [184], *Rana japonica* [200-202, 214, 216], *Rana temporaria* [216], *Rana pipiens* [216], *Xenopus laevis* [216], *Rana brevipoda porosa* [216], рыбы *Cnesterodon decemmaculatus* [86], *Brachydanio rerio* [109, 187, 198], *Danio rerio* [189], *Lepomis macrochirus* [109, 110, 182, 194, 197, 218], *Oryzias latipes* [109, 110, 196, 197], *Leuciscus idus* [109, 182, 197], *Carassius auratus* [197], *Ictalurus* sp. [182], членистоногие *Daphnia carinata* [80], *Daphnia pulex* [182, 190], *Gammarus pulex* [193], водоросли *Pseudokirchneriella subcapitata* [189, 192, 198, 199, 218], *Desmodesmus subspicatus* [192], *Scenedesmus subspicatus* [182, 198], *Scenedesmus obliquus* [79, 80, 205-209], *Scenedesmus vacuolatus* [210], *Selenastrum capricornutum* [182], *Chlorella pyrenoidosa* [80], *Chlorella fusca* [211], простейшие *Entosiphon sulcantum* [191], *Uronema parduczi* [191], *Chilomonas paramecium* [191], грибы *Aspergillus nidulans* [195], дрожжи *Saccharomyces cerevisiae* [80, 203, 204], бактерии *Shk1* [212, 213], *Burkholderia species Rasc c2* [220], *Pseudomonas fluorescens* [220, 228], *Escherichia coli* [189, 224], *Sinorhizobium meliloti* [227, 230], *Pseudomonas putida* [214], бактерии из речной воды [79, 215], инфузории *Tetrahymena thermophila* [217], пчелы [219, 225, 229], огурцы *Cucumis sativus* [221-223], комнатные мухи [226], птицы *Quail* [229], черви *Tubifex tubifex* [231-233], микроорганизмы активированного ила [234]. Также имеется небольшое количество публикаций, связанных с токсичностью

химических соединений по отношению к человеку [227, 235-237]. Обобщенный нами результат анализа частоты встречаемости различных организмов при исследовании острой токсичности представлен на рис. 2.

В качестве количественных мер токсичности исследуемых соединений используются LC_{50} (средняя летальная концентрация, вызывающая гибель 50% организмов), EC_{50} (средняя эффективная концентрация, при действии которой на объект развивается эффект, равный 50% от максимально возможного эффекта) и LD_{50} (средняя летальная доза, вызывающая гибель 50% организмов). Дополнительными характеристиками, связанными с мерами токсичности, являются время экспозиции и способ введения вещества. Так, в случае рыбы *Pimephales promelas*, в качестве меры токсичности выступает концентрация LC_{50} , со временем экспозиции 96 часов [238]. Для *Tetrahymena pyriformis* мерой токсичности служит концентрация EC_{50} , вызывающая ингибирование роста инфузорий с инкубационным периодом 40 часов [239]. Токсичность химических соединений по отношению к крысам при ингаляционном способе введения вещества выражается в виде LC_{50} , (экспозиция 4 часа) [240], а при внутривенном введении в виде LD_{50} (24 часа) [241].

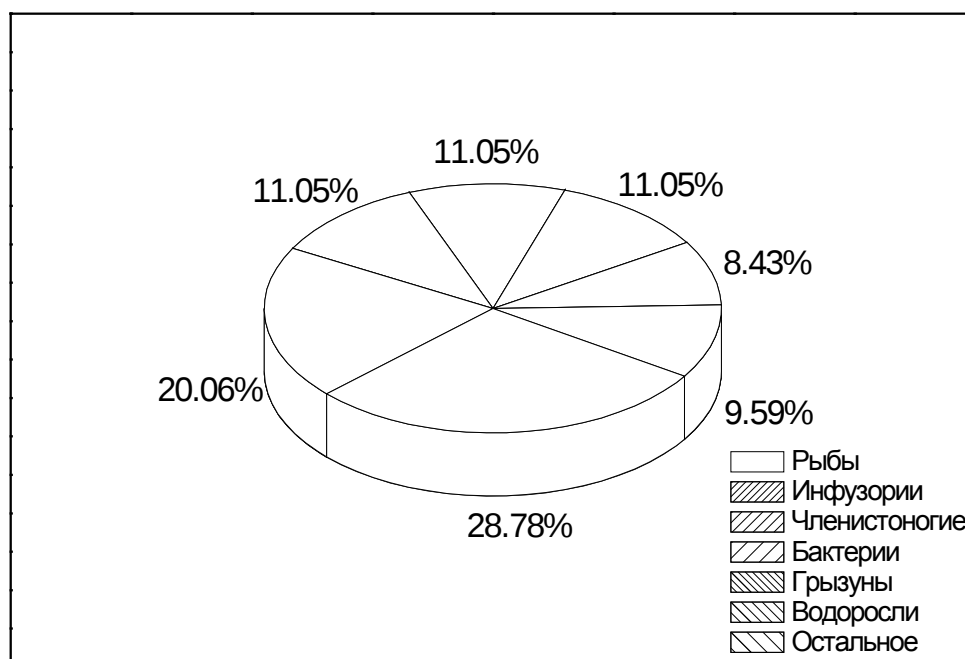


Рис. 2. Частота встречаемости организмов при исследовании острой токсичности органических соединений

Источниками токсикологических данных могут выступать различные базы данных [6]. БД ECOTOX [312] по состоянию на 15.06.11 включала в себя 786447 записей, 9783 водных и земных организмов и 9727 химических соединений. БД ESIS [242] содержит свыше 38000 записей по острой токсичности и свыше 3000 записей по хронической токсичности. По состоянию на 05.07.11 БД OECD NPV [243] содержала данные о 4951 химических соединений, производящихся в больших масштабах. TOXNET является кластером различных БД, содержащих данные по токсичности нескольких тысяч соединений [244]. В БД N-CLASS включены сведения о 7897 веществах [245]. Информация о влиянии химических соединений на окружающую среду и здоровье человека отражена в RISKLINE [246]. Обширная база данных RTECS [313] содержит сведения о приблизительно 169000 соединений. База данных ACCELRYIS [247] включает в себя полное содержимое RTECS и другие БД по токсичности. Специализированная БД

TETRATOX [248] является коллекцией данных по острой токсичности (EC_{50} , 40 час.) более чем 2400 химических соединений по отношению к инфузории *Tetrahymena pyriformis*.

В литературе также публикуются данные о токсичности достаточно больших рядов соединений по отношению к различным организмам. Так, в работе [238] приведены данные по острой токсичности (LC_{50} , 96 час.) 617 соединений, классифицированных по 9 MOA, по отношению к рыбе *Pimephales promelas*. В качестве другого биологического объекта фигурирует рыба *Poecilia reticulata* [16], публикуются сведения о токсичности 293 соединений с 4 MOA. Для другой рыбы *Oncorhynchus mykiss* приводятся сведения о токсичности 282 пестицидов [117, 124]. В работе [144] приведены величины токсичности (LC_{50} , 48 час.) для 300 различных соединений по отношению к *Daphnia magna*. Для этого организма известны также данные по токсичности 262 пестицидов [150].

1.3. Дескрипторы

Для описания структуры химических соединений используются различные дескрипторы, которые могут быть разбиты на три группы: 2D, 3D и физико-химические [11]. 2D дескрипторы рассчитываются на основе 2D структуры, при этом широко используется теория графов. 3D дескрипторы требуют знания геометрической и электронной структуры соединений. Физико-химические дескрипторы характеризуют свойства целой молекулы в виде простого числа. Дескрипторы также могут быть классифицированы по их физической природе или методу расчета на следующие типы: квантово-химические, топологические, электронные, фрагментные, конституционные, геометрические, электростатические, термодинамические, бинарные и другие. Для их расчета применяются такие компьютерные программы, как DRAGON [315], MOPAC [249], TSAR [250], HYPERCHEM [321], CODESSA [251], SYBYL [252], OASIS [253], GAUSSIAN [254], CERIUС [255]. Число рассчитываемых дескрипторов колеблется от 1 до 5000.

В качестве наиболее часто используемого дескриптора выступает $\log K_{ow}$ (коэффициент распределения вещества в системе н-октанол-вода). Он встречается практически во всех анализируемых работах. Так, в двадцати шести из двадцати восьми найденных публикаций, в которых для построения QSAR моделей используется только один дескриптор, им оказывается $\log K_{ow}$. В одиннадцати работах для анализа токсичности используется два дескриптора, во всех случаях один из них – $\log K_{ow}$. Для расчета $\log K_{ow}$ используются такие компьютерные программы, как CLOGP [256], KOWWIN [257].

Следует отметить, что во многих публикациях [35, 50, 67, 72, 183, 208, 223, 258-263, 265, 281, 282] подчеркивается важность водородной связи при проявлении токсичности. Так, при исследовании токсичности 50000 соединений [259], которые были разбиты на три группы: яды, лекарства и природные соединения, была обнаружена положительная корреляция между токсичностью и числом протоноакцепторов и протонодоноров в молекулах. При этом было установлено, что природные соединения, лекарства и яды с низкой и средней токсичностью имеют одинаковое количество протоноакцепторных и протонодонорных центров, число которых колеблется между тремя и шестью для протоноакцепторов и между нулем и двумя для протонодоноров. Полученный результат хорошо согласуется с известным «правилом пяти» Липинского [260], согласно которому лекарства с хорошими ADME свойствами должны иметь не более пяти Н-доноров и десяти Н-акцепторов.

В работе [258] для характеристики способности замещенных анилинов к образованию Н-связи используется комбинация трех дескрипторов: E_{LUMO} , E_{HOMO} и Q^+ . Полученные результаты показывают, что токсичность возрастает с возрастанием протонодонорной способности группы NH_2 и поддерживают гипотезу о том, что механизм токсического действия связан с образованием водородных связей между аминогруппами анилинов и полярными группами клеточных мембран.

Исследуя новый подход для предсказания механизмов токсического действия фенолов, авторы публикации [261] использовали в качестве одного из девяти дескрипторов число Н-доноров в молекуле. Ими была установлена важность этого дескриптора при анализе фенолов с МОА полярный наркоз. Аналогичный результат получен при классификации фенолов на три группы с использованием дискриминантного анализа [262]. Для описания структуры молекул использовалось четыре дескриптора, два из которых характеризуют водородную связь. При использовании двухшаговой схемы классификации 220 фенолов по четырем МОА [72] также отмечается, что в случае полярных наркотиков необходимой комбинацией дескрипторов является $\log K_{ow}$ (гидрофобность) и N_{Hdon} (число Н-доноров). Подобный результат получен при классификации этих же фенолов с применением пошагового LDA [35] и регрессионных классификационных методов [67]. Авторы работы [50], используя методы ANN и LR, отмечают улучшение дискриминации между фенолами: проэлектрофилами и полярными наркотиками при введении в качестве дополнительного дескриптора N_{Hdon} . Успешная дискриминация между фенолами с двумя МОА (полярные наркотики и разобщающие агенты) была проведена с помощью SVM на основе семи дескрипторов [265]. Характеристиками водородной связи были N_{Hdon} , N_{Hacc} , E_{HOMO} . Всего было получено восемь моделей классификации, насчитывающих от 1 до 7 дескрипторов. В семи из них, за исключением однопараметровой модели с $\log K_{ow}$ в качестве переменной, фигурировали параметры Н-связи.

Важными факторами для проявления токсичности неполярных и полярных наркотиков являются гидрофобность, выражаемая через $\log K_{ow}$, и способность к образованию водородной связи [281, 282]. Для количественного выражения последнего свойства использован набор из четырех дескрипторов: E_{LUMO} , E_{HOMO} , Q^+ (наибольший положительный заряд на атоме водорода), Q^- (наибольший отрицательный заряд на неводородном атоме). По мнению авторов, E_{LUMO} и E_{HOMO} описывают ковалентный вклад в водородную связь для Н-доноров и Н-акцепторов соответственно, а Q^+ , Q^- отражают ионный вклад.

Полученные регрессионные зависимости отражают вклады дескрипторов в токсичность исследованных соединений.

В работе [263] на основе нелинейных дискриминантных функций получена хорошая классификационная модель для 150 различных соединений, относящихся к двум МОА: неполярный и полярный наркоз. Дескрипторное пространство состояло из пяти вышеприведенных переменных: $\log K_{ow}$, E_{LUMO} , E_{HOMO} , Q^+ , Q^- .

Авторы публикации [183] с использованием QSAR исследовали токсичность 104-х алкалоидов. Ими была выявлена группа из семи дескрипторов, на основе которых исследуемые соединения были разделены на «лекарства» и «не лекарства». Один из этих дескрипторов являлся суммой протоноакцепторных атомов, а другой – суммой протонодонорных групп в молекуле.

На основе дескрипторов, входящих в состав регрессионного уравнения, описывающего токсичность фенолов по отношению к *Cucumis sativus*, делается вывод о важности гидрофобности, электрофильности и способности к образованию Н-связи для проявления токсичности [223]. В качестве количественной меры способности молекулы действовать как протоноакцептор использовалась величина максимального отрицательного парциального атомного заряда. Также о важности таких дескрипторов, как поляризуемость, основность водородной связи и молекулярный объем для описания токсичности замещенных фенолов свидетельствуют данные работы [208].

1.4. Кластеризация молекул. Предобработка данных

Одним из предположений, положенных в основу методов QSAR, является гипотеза о том, что подобные вещества обладают одинаковым механизмом действия [2]. В соответствии с этим принципом для конструирования локальных количественных моделей, связывающих токсичность химических соединений с их структурными характеристиками, необходима

предварительная группировка (кластеризация) молекул [96]. При этом используются различные подходы для кластеризации.

Одним из подходов является формирование выборок соединений из одного химического класса. В литературе имеются сведения о моделях QSAR для рядов соединений с различными функциональными группировками: фенолы [33, 34, 36, 44, 50, 53, 65, 67, 68, 70, 72, 143, 201, 206, 207, 214, 221-223, 234, 261, 265], анилины [143, 189, 190, 206, 207, 234], спирты [155, 168-170], кетоны [29, 170], альдегиды [29, 48, 55, 57, 59, 76, 101, 266], нитроароматические соединения [40, 49, 69, 78, 88, 177, 190, 204, 210, 217, 267], эфиры [163], сложные эфиры [60, 151], амиды [176], азотсодержащие гетероциклические соединения [126]. Число используемых классов химических соединений для оценки токсичности может исчисляться десятками. Так в системе KATE [196] используется 44 класса химических соединений.

Ведущий подход для группировки молекул основан на использовании концепции типов токсического действия (МОА). В литературе описываются различные подходы для установления МОА [19]: субструктурный анализ молекул [268, 269], концентрационные изменения токсичности в смесях [270], высококачественные модели QSAR, основанные на экспериментальных данных [271], анализ объемных долей [272], комбинации статистических методов и различных параметров [273], синдром острой токсичности у рыб (FATS) [274]. Например, в схеме Верхаара (Verhaar) [268], используется четыре класса соединений. Класс 1 составляют инертные химические соединения, класс 2 – это менее инертные соединения, реактивные соединения образуют 3 класс и специфические соединения входят в состав 4 класса. Считается, что соединения 1 класса не взаимодействуют со специфическими рецепторами в организме и относятся к МОА неполярный наркоз [6]. Менее инертные соединения показывают несколько большую токсичность по сравнению с соединениями 1 класса, обладают повышенной протонодонорной способностью и относятся к МОА полярный наркоз. Класс 3 включает различные реактивные соединения, обладающие повышенной токсичностью, обусловленной неселективным

взаимодействием с биомолекулами, или связанной с образованием более токсичных метаболитов. Класс 4 наиболее часто связывают с соединениями, действующими как ингибиторы AChE, или оказывающими влияние на центральную нервную систему.

Для классификации химических соединений также используется схема, предложенная в работе [238]. Она включает в себя группировку молекул с применением восьми МОА: наркоз I (неполярный наркоз), наркоз II (полярный наркоз), наркоз III (эфирный наркоз), разобщение окислительного фосфорилирования, ингибирование дыхания, электрофильная или проэлектрофильная реактивность, ингибирование AChE, действие на центральную нервную систему.

Большое внимание исследователи посвящают изучению токсичности химических соединений с типом токсического действия наркоз [62, 86, 93, 134, 154]. Имеются публикации, в которых получены QSAR модели для выборок соединений с одним МОА: неполярный наркоз [29, 54, 155, 239, 240], полярный наркоз [33, 34, 143, 189, 206], реактивность [31, 60, 88]. Ряд работ связан с изучением двух МОА: неполярный наркоз и полярный наркоз [18, 30, 128, 132, 191, 213, 216, 263, 310], наркоз и реактивность [19], неполярный наркоз и реактивность [108], полярный наркоз и электрофильность [64, 208], полярный наркоз и реактивность [20, 221, 223]. Также известно большое количество работ, в которых химические соединения характеризуются тремя и более МОА [16, 17, 21-23, 25, 27-29, 31, 35, 43-45, 50, 55-57, 65, 67, 68, 70-73, 84, 85, 90, 95, 97, 102, 106, 107, 112, 118, 121, 140, 193, 201, 204, 215, 230, 232, 261, 262, 264, 275-280].

В литературе описаны и другие способы предварительной группировки химических соединений. Так, в работе [122] для проведения кластеризации 164 пестицидов из 7 химических классов в пространстве из 174 молекулярных дескрипторов используется метод GA. В качестве другого способа классификации химических соединений применяется разбиение интервалов изменения токсичности LC_{50} или LD_{50} на три [166, 229] или четыре класса

[100]. В публикации [192] используется предварительная кластеризация соединений на 4 группы на основе отношения LC_{50}/EC_{50} . Авторами работы [94] проведена группировка 568 органических молекул на 7 классов на основе четырех дескрипторов: число ароматических колец, число алифатических колец, $\log K_{ow}$, максимальный порядок связи атома углерода.

При разработке глобальных моделей токсичности химических соединений и отсутствии информации о механизмах их токсического действия проводится группировка молекул на основе различных количественных мер структурного сходства: коэффициента Танимото [180], косинусного коэффициента [178], Эвклидова расстояния [178]. Также для формирования кластеров находит применение метод HCA [178].

Следует отметить, что при конструировании глобальных QSAR моделей токсичности предварительная кластеризация проводится не всегда. Для получения таких моделей используются методы ANN [42, 89, 98, 103, 105, 117, 311], PLS [39, 311], MLR [42, 89, 109, 148, 311], SVM [98, 311], KNN [178, 311]. При этом для уменьшения размерности пространства и выделения информативных дескрипторов, как правило, используется отбор дескрипторов с помощью GA [105, 109, 117, 148] и PCA [105].

Обычно следующим шагом после выделения кластера молекул и связанного с ним набора дескрипторов является предобработка данных. Целью этого этапа является подготовка данных к дальнейшему использованию. Предобработка может включать в себя такие процедуры, как пересчет единиц измерения токсичности из весовых единиц в молярные единицы [89, 101, 152, 177], логарифмирование величин токсичности [89, 94, 95, 101, 149, 152, 177], стандартизация независимых переменных, например, центрирование или шкалирование [95, 96], которые часто объединяются в одну процедуру: автошкалирование [69, 94, 264].

1.5. Отбор дескрипторов

Отбор дескрипторов является важным, ключевым шагом при построении моделей [158, 159, 264]. Иногда селекция дескрипторов не проводится. Обычно это связано с наличием априорного знания о факторах, влияющих на токсичность, и небольшого числа переменных [17, 157]. В случае небольшого числа используемых дескрипторов возможен полный перебор их комбинаций [92, 190]. Однако, гораздо чаще, приходится иметь дело с большим количеством дескрипторов. При этом с увеличением их числа очень быстро возрастает число возможных моделей как $N!/(N-I)!I!$ [92], где N - общее число дескрипторов, I – число дескрипторов используемых в модели. Полный перебор всех комбинаций переменных в этом случае проводится редко [165, 227], ему может предшествовать предварительный анализ корреляции переменных с токсичностью [57].

Чтобы избежать экспоненциального взрыва, связанного с перебором всех возможных комбинаций переменных, применяется несколько методов. Их условно можно разбить на две группы. К первой группе относятся методы отбора дескрипторов выполняемые как отдельные процедуры, предшествующие созданию количественных моделей токсичности. Одними из самых простых подходов являются: удаление независимых переменных, которые имеют постоянное значение в 100% случаев [166], более чем в 95% случаев [230], или более чем в 90% случаев [95, 96, 98, 159], или у которых относительное стандартное отклонение меньше 0.05 [98], или которые имеют коэффициент корреляции с токсичностью $r^2 < 0.1$ [94]. Также находит применение анализ корреляционной матрицы дескрипторов [88, 89, 94-96, 98, 101, 113, 137, 140, 149, 152, 159, 165, 166, 176, 177, 185, 208, 209, 223, 230]. При этом, обычно, отбрасывается один из дескрипторов в паре, если коэффициент парной корреляции превышает некоторое пороговое значение, например, $r > 0.4$ [185], $r > 0.5$ [49], $r > 0.8$ [88, 89, 113], $r > 0.85$ [95, 96], $r > 0.9$ [98, 159], $r > 0.95$ [230], $r > 0.98$ [165], $r^2 > 0.6$ [94, 101, 177], $r^2 > 0.9$ [264] или $r^2 > 0.95$ [166]. Кроме того, для

выбора дескрипторов используются генетический алгоритм [48, 76, 86, 91, 97, 100, 105, 109, 117, 122, 148, 149, 151, 152, 175, 177, 178, 179, 181, 183, 203, 211, 214, 222, 225, 229, 264, 311], анализ главных компонент [37, 47, 78, 105, 114, 140, 141, 203, 222, 275], факторный анализ [75, 160, 202, 236], метод ближайшего соседа [17, 311], кросс-валидация [98], HSA алгоритм [229].

Ко второй группе относятся подходы, в которых отбор дескрипторов и построение количественных QSAR моделей происходит одновременно. Сюда можно отнести различные пошаговые методы, в которых на каждом шаге происходит включение/исключение переменной в модель на основе значения некоторого критерия. После получения ряда моделей на основе статистических величин выбирается «лучшая» из них. Подобного рода процедуры связаны с использованием таких методов как множественная линейная регрессия [16, 17, 26, 27, 36, 40, 42, 53, 61, 64, 70, 73, 75, 76, 79, 88, 101, 118, 121, 159, 140, 149, 152, 160, 164, 184, 185, 200, 203, 208, 209, 222, 236, 311], проекции на скрытые структуры [157], искусственные нейронные сети [42, 73], дискриминантный анализ [19, 35, 72, 264]. Следует отметить также появление новых методов, например, программирование экспрессии генов (GEP) [158], сочетающее в себе одновременный выбор переменных и расчет нелинейных моделей.

В литературе имеются описания оригинальных методов, связанных с селекцией переменных. Среди них можно отметить: анализ графика зависимости r^2 от числа дескрипторов в регрессионной модели [101], TS алгоритм, представляющий из себя итерационную процедуру для решения оптимизационных проблем [159], оптимизация корреляционных весов дескрипторов [104, 177], выбор дескрипторов для нелинейных систем с помощью подхода Shuffling-ANFIS [74], эвристический CfsSubset алгоритм, основанный на анализе корреляций переменных между собой и корреляции переменных с целевым свойством [261], метод Монте-Карло, симулирующий физический процесс отжига [98], эвристический подход НМ, сочетающий в себе контроль коллинеарности дескрипторов и быстрый перебор регрессионных моделей [125, 161], метод RFE, показавший свою

эффективность в сочетании с классификационными процедурами [41], процедура RM [43], которая по своей эффективности приближается к полному перебору всех комбинаций дескрипторов в регрессии, но требует значительно меньше времени.

1.6. Статистические методы моделирования

Для создания количественных QSAR моделей острой токсичности химических соединений используются различные статистические методы: множественный регрессионный анализ (MLR) [16-18, 22, 23, 25-34, 36, 38, 40, 42, 43, 45-47, 49, 53-55, 57, 60-64, 66, 67, 69, 70, 73, 75, 76, 78-81, 85-89, 91, 92, 94, 97, 101, 104, 106-110, 112-121, 125, 126, 130, 132, 133, 135-137, 139-141, 143, 144, 146-153, 155, 157, 159, 160, 162-165, 169, 170, 175-177, 181-191, 193, 195-197, 199-204, 206-211, 213, 215-217, 219, 221-223, 226, 227, 230, 232-237, 239-241, 258, 277-280, 310, 311], проекции на скрытые структуры (PLS) [23, 27, 32, 37, 39, 48, 57, 71, 75, 76, 93, 95, 96, 117, 121, 131, 142, 149, 152, 156, 157, 160, 172, 179, 203, 208, 214, 218, 222, 230, 236, 264, 275, 311], искусственные нейронные сети (ANN) [12, 13, 34, 37, 38, 41, 42, 44, 50, 71, 73, 74, 83, 89, 98, 103, 105, 106, 117, 125, 158, 161, 165, 179, 194, 225, 261, 311], опорные вектора (SVM) [20, 34, 41, 44, 56, 77, 90, 98, 102, 125, 158, 261, 265, 311], дискриминантный анализ (DA) [19, 35, 44, 67, 72, 121, 262-264, 275, 276], метод k-ого ближайшего соседа (KNN) [41, 77, 166, 173, 178, 261, 311], логистическая регрессия (LR) [19, 41, 50, 65, 67, 72, 264], метод случайного леса (RF) [77, 166, 178], дерево решений (DT) [41, 68, 77], адаптивное нечеткое распределение (AFP) [100, 229], программирование экспрессии генов (GEP) [158]. Рассчитанные нами частоты использования статистических методов исследования представлено на рис. 3.

Из представленных данных следует, что наиболее часто при создании QSAR моделей острой токсичности химических соединений используется метод MLR. Достаточно часто применяются такие подходы как PLS и ANN.

Другие методы используются реже. В связи с полученными данными представляет интерес сравнение свойств методов MLR, PLS и ANN (табл. 1). Как следует из данных этой таблицы, в одних случаях преимущество имеют одни методы, в других случаях – другие методы.

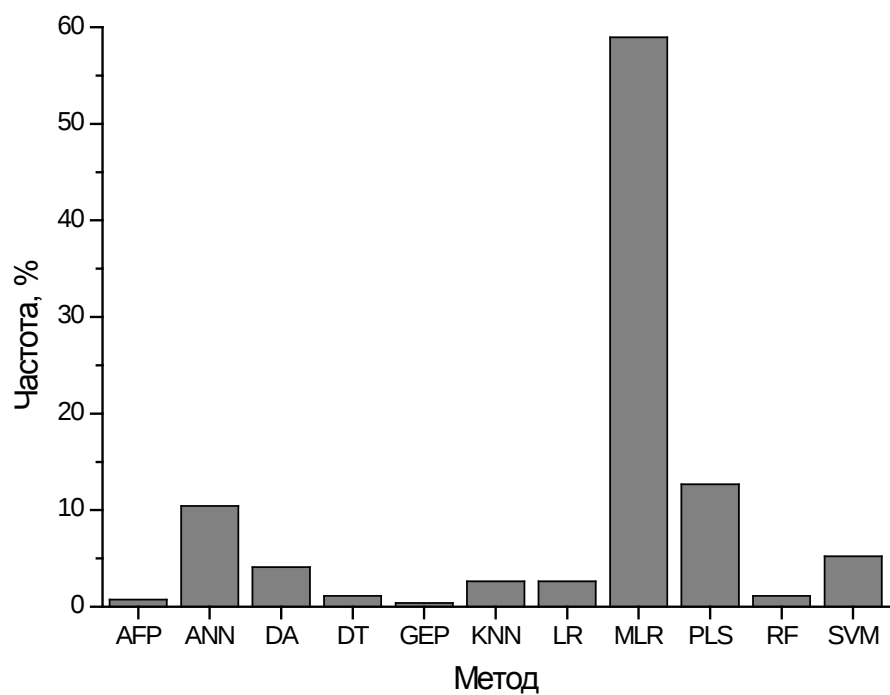


Рис. 3. Частота использования статистических методов при исследовании острой токсичности органических соединений

Таблица 1

Сравнение свойств токсикологических моделей QSAR, основанных на MLR, PLS и ANN [27]

Свойство	MLR	PLS	ANN
Предсказательность	Хорошая для линейных моделей, плохая для остальных	Превосходная для линейных и хорошая для нелинейных моделей	Превосходная для линейных и нелинейных моделей
Понятность	Превосходная: модели ясно демонстрируются и легко описываются	Умеренная: модели могут быть интерпретированы, но это может быть затруднительно в случае большого числа параметров	Плохая: точность модели и влияние параметров остаются в значительной степени неизвестными
Полезность: переносимость	Превосходная: предсказания могут быть сделаны на основе знания соответствующих физико-химических свойств	Умеренная: предсказания могут быть сделаны из новых размерностей, но требуется трансформация оригинальных переменных	Плохая: модели должны быть перетренированы до использования
Полезность: добавление новых данных	Хорошая: модели легко пересчитать	Умеренная: модели нуждаются в пересчете	Плохая: пересчет требует много времени
Механистическая значимость	Хорошая: влияние индивидуальных физико-химических дескрипторов может быть оценено	Умеренная: возможно применение механистической интерпретации к новым размерностям	Плохая: влияние индивидуальных дескрипторов редко подчеркивается

Продолжение табл. 1

Свойство	MLR	PLS	ANN
Возможность иметь дело с многомерными проблемами	Плохая-умеренная: многие проблемы обнаруживаются в связи с коллинеарностью переменных и отношением числа наблюдений к числу переменных	Превосходная: подходящий метод чтобы иметь дело с большими матрицами коллинеарных данных	Превосходная: имеется возможность разрабатывать модели с большим числом переменных
Требования к данным	Хорошие: могут быть разработаны модели как с малым, так и с большим числом данных	Хорошие: могут быть разработаны модели как с малым, так и с большим числом данных	Умеренные: обычно требуется большой ряд данных
Легкость использования и доступность	Хорошая: регрессионный анализ является простым статистическим методом, имеющимся во многих статистических пакетах программ	Умеренная: PLS менее доступен, необходим ряд расчетов, чтобы сделать предсказание	Плохая: требуются специалист по компьютерным программам и знания по их использованию

1.7. Валидация моделей

Важным моментом при QSAR моделировании является валидация моделей [283-285]. Первым шагом на этом пути служит проверка полученных результатов на выбросы. Известны выбросы трех видов: Y-, X- и X/Y-выбросы [218]. Y-выбросы связаны со значениями активности (токсичности) для обучающей и тестовой выборок. X-выбросы связаны с дескрипторами, которые не попадают в область допустимых значений обучающего ряда. X/Y-выбросы связаны с химическими соединениями тестовой выборки, у которых связь между зависимыми и независимыми переменными не такая, как в обучающей выборке. Анализ на выбросы может выполняться как в виде отдельной процедуры, так и быть частью другой стороны валидации: определения допустимой области применения модели. В первом случае, чаще всего, речь идет об Y-выбросах. Один из самых простых подходов для их анализа состоит в изучении стандартизованных (в единицах стандартного отклонения s) остатков т.е. разностей между экспериментальными и рассчитанными величинами активности (токсичности). Точки, отклоняющиеся более чем, например, на $2s$ [75] или $3s$ [25, 39], отбрасываются и модель пересчитывается заново. Более усложненной процедурой является графический анализ нормальных вероятностей стандартизованных остатков [287].

В соответствии с OECD принципами [286] можно выделить три аспекта валидации: описательная способность, предсказательная способность и область применимости. Обычно для характеристики описательной способности моделей используются следующие величины: число точек (n), коэффициент линейной корреляции (r), стандартное отклонение (s), критерий Фишера (F). Их совокупность является достаточной для того, чтобы ответить на вопрос, насколько хорошо предлагаемая модель воспроизводит экспериментальные данные. Наиболее популярной характеристикой является r . Обычно, модель считается удовлетворительной, если $r^2 > 0.6$ [284].

Для оценки предсказательной способности моделей используется два подхода: внешняя валидация и внутренняя валидация. При внешней валидации доступные данные по активности разбиваются на два ряда. Первый ряд (обучающая выборка) служит для создания модели. Второй ряд (тестовая выборка) служит для оценки предсказательной способности. При этом основные различия в подходах разных авторов заключаются в выборе метода разделения данных и объемах выборок. Для характеристики предсказательной способности при внешней валидации служат соответствующие коэффициенты r^2_{ext} и q^2_{ext} [33].

Для генерации обучающих и тестовых выборок используются следующие методы: случайное разделение [45, 65, 68, 85, 93, 96, 103, 105, 112, 117, 119, 120, 137, 152, 156, 158, 159, 161, 165, 166, 177, 218, 225, 311], кластеризация К-средними [37, 66, 75, 149, 236], сферическое исключение [173], упорядочивание по токсичности [17, 35, 39, 48, 72, 157, 165, 190, 197], SMD алгоритм [218], разные источники данных [91, 98, 113, 179, 241], комбинация методов НСА и РСА [33], искусственная нейронная сеть Кохонена [97, 165], анализ химического пространства [41], неоднородность данных [25].

Объемы тестовой выборки (% от всей выборки) изменяются в больших пределах: 4% [98], 5% [113], 7% [194], 10% [50, 85, 103, 112, 187], 11% [225], 13% [119, 137], 14% [34, 156, 241], 15% [38, 120, 179], 16% [181], 19% [117, 125, 219, 227], 20% [17, 37, 43, 66, 71, 73, 74, 89, 95, 125, 165, 166, 190, 234], 21% [70], 23% [158, 91], 24% [115], 25% [48, 75, 76, 96, 149, 236], 27% [170], 28% [91], 30% [65, 68, 105], 32% [159, 164], 33% [41, 42, 93, 165, 186], 34% [311], 35% [90], 36% [104, 148], 38% [133], 39% [109], 50% [35, 39, 72, 152, 157, 177, 197, 218], 51% [33, 59, 266], 53% [90, 178], 83% [239], 92% [50].

Как следует из представленного материала, наиболее часто используемым подходом для получения тестовых данных является случайное генерирование последовательностей химических соединений. Объем тестового ряда составляет, чаще всего, 20% от общего размера выборки.

В качестве распространенных подходов для внутренней валидации моделей служат: кросс-валидация с выбором по одному (LOO), кросс-валидация с выбором по N (LNO), бутстреп, Y-рандомизация [33].

LOO является простой и распространенной процедурой используемой для валидации моделей токсичности химических соединений [16-19, 25-28, 34, 37, 39, 40, 42-44, 46-49, 54-57, 61-64, 66, 73-76, 78, 80, 86, 89, 90, 92-95, 97, 101, 109, 114, 121, 122, 125, 131, 133, 141, 142, 148, 149, 152, 156, 157, 160, 161, 165, 173, 176, 180, 183, 184, 186, 187, 190, 195, 196, 199, 200, 203, 208, 209, 211, 218, 222, 226, 227, 230, 232, 233, 235-237, 258, 278, 311]. Она заключается в исключении каждого образца (химического соединения) однажды, конструировании новой модели без этого образца и предсказании на основе этой модели значения зависимой переменной (токсичности) для исключенного образца. Иными словами, для обучающей выборки, состоящей из M образцов, LOO выполняется M раз путем удаления первого, второго и т.д. образца давая в итоге M предсказанных величин. Для оценки предсказательной способности служит величина q^2 . QSAR модель считается адекватной при условии, что $r^2 > 0.6$, $q^2 > 0.5$, $(r^2 - q^2) < (0.2 \dots 0.3)$ [33].

Для тестирования разработанных моделей QSAR также используется LNO кросс-валидация [33]. При ее выполнении обучающий ряд из M образцов делится на последовательные блоки, состоящие из N образцов. Таким образом, число блоков равно целой части отношения M/N плюс неполный комплект, который обычно образует последний блок. LNO кросс-валидация основана на тех же принципах, что и LOO: каждый блок удаляется однажды, новая модель создается без него, на основе этой модели предсказываются значения зависимой переменной для блока. LNO выполняется для N=2, 3, и т.д., при этом коэффициент корреляции q^2_{LNO} рассчитывается также, как и q^2 . Так как величина q^2_{LNO} , в отличие от q^2 , зависит от порядка следования образцов, имеет значение метод отбора образцов в блоки. При валидации количественных моделей токсичности используется LNO с блоками различной длины: N=2 [90], N=3 [90, 101], N=5 [41, 50, 67, 70, 77, 90, 91, 102, 166, 173, 264, 265, 278, 279],

$N=10$ [74, 90, 229, 232, 279], $N=30$ [74], $N=50$ [279], $N=53$ [74], $N=10\%$ [20, 113], $N=30\%$ [211] и различные методы формирования блоков: упорядочивание по индексу Танимото [77], случайный отбор [50, 67, 74, 90, 91, 166, 173, 211, 265, 232], упорядочивание по токсичности [101, 264]. Как следует из представленных данных, наиболее часто используется LNO с $N=5$ и случайный способ образования блоков.

Для тестирования QSAR моделей острой токсичности химических соединений также используется Y-рандомизация [26, 48, 74, 97, 152, 211]. Целью этого метода анализа является оценка случайной корреляции между дескрипторами и зависимой переменной. Процедура проводится путем повторения ряда итераций, при которых матрица независимых переменных (дескрипторов) остается неизменной, а вектор зависимой переменной (активность) меняется случайным образом. Модели, полученные при таких условиях, обладают плохими характеристиками и не имеют физического смысла. Для количественной оценки Y-рандомизации служат величины r^2_{rand} и q^2_{rand} . При проведении Y-рандомизации необходимо ответить на два вопроса: как анализировать результаты расчетов и сколько делать итераций [33]? Один из простых подходов заключается в сравнении величин r^2_{rand} и q^2_{rand} :

$q^2_{\text{rand}} < 0.2$ и $r^2_{\text{rand}} < 0.2$ случайная корреляция отсутствует,
любое q^2_{rand} и $0.2 < r^2_{\text{rand}} < 0.3$ слабая случайная корреляция,
любое q^2_{rand} и $0.3 < r^2_{\text{rand}} < 0.4$ терпимая случайная корреляция,
любое q^2_{rand} и $r^2_{\text{rand}} > 0.4$ случайная корреляция.

Другой подход [91] состоит в сравнении величин r^2 тестируемой модели, среднего значения r^2_{rand} и стандартного отклонения s для набора r^2_{rand} :

$(r^2 - r^2_{\text{rand}}) \geq 2.3 s$ вероятность случайной корреляции находится на уровне 1%,

$(r^2 - r^2_{\text{rand}}) \geq 3 s$ вероятность случайной корреляции находится на уровне 0.1%.

Также для изучения случайной корреляции применяются графические методы анализа [117].

Что касается количества итераций, то в литературе можно найти различные ответы на этот вопрос: используется 10 [33, 41, 173], 25 [33], 30 [98], 100 [39, 91, 113, 117], 500 [165] итераций.

Бутстреп является редко используемым методом валидации моделей острой токсичности соединений. При применении этого подхода исходные данные случайным образом расщепляются на обучающую и тестовую выборку несколько раз. В отличие от LOO и LNO, где каждое соединение используется один раз, в бутстрепе соединение может использоваться в процедуре один раз, много раз или вообще не встречаться. В качестве характеристик бутстреп служат величины r^2_{bstr} и q^2_{bstr} для обучающей и тестовой выборок соответственно. При тестировании моделей с помощью этого метода сталкиваются с двумя проблемами: формирование выборок и число итераций. В литературе последнего десятилетия мы нашли небольшое количество работ, в которых используется бутстреп для оценки моделей острой токсичности. При этом в качестве методов образования обучающих и тестовых рядов химических соединений использованы: случайный выбор из полного набора данных [33], выбор из HSA кластеров [33] и выбор из PCA кластеров [33]. Число итераций составляло: 10 [33], 25 [33], 5000 [97, 165].

Необходимым элементом валидации полученных QSAR моделей является определение области допустимых значений (AD). В широком плане AD модели означает физико-химическое, структурное или биологическое пространство, информацию о котором предоставляет обучающая выборка, и для которого действительно предсказание активности для новых соединений [152]. В узком плане AD можно интерпретировать как область химических структур, для которых применима рассматриваемая модель. Известны различные статистические подходы для оценки AD: интервальный, геометрический, дистанционный, плотность вероятности [288].

В качестве простого метода оценки AD используются интервалы индивидуальных дескрипторов. В случае n переменных область допустимых значений выглядит как n -мерный гиперпрямоугольник со сторонами

параллельными осям координат. Ограничения этого метода связаны с наличием неопределенного пустого пространства и отсутствием учета корреляции между дескрипторами. В какой-то степени эти недостатки могут быть преодолены при использовании вместо индивидуальных дескрипторов их линейных комбинаций: главных компонент.

Для геометрического определения AD n -мерного ряда соединений применяется минимальная выпуклая оболочка, которая содержит все соединения из обучающего ряда. При расчете такой оболочки появляются определенные проблемы, связанные с увеличением числа точек и размерности пространства. К недостаткам метода можно отнести отсутствие идентификации пустого пространства.

В случае использования дистанционного подхода рассчитывается расстояние от интересующей точки до выборки данных. При этом необходимо учитывать два момента: используемая метрика и граничное значение расстояния. Популярными дистанционными мерами являются Эвклидово расстояние и расстояние Махаланобиса. Также в качестве меры расстояния используются диагональные элементы матрицы проекций на пространство регрессоров (Hat matrix), так называемые левэриджи (leverage) h . При использовании этой метрики в качестве граничного значения выступает величина $h^* = 3p/n$, где n – число соединений в обучающей выборке, p – число переменных в модели плюс 1. Также для анализа AD моделей используется график Вильямса, представляющий из себя проекции соединений на плоскость, образованную величинами h и стандартизованными остатками.

Функция плотности вероятности может быть оценена как параметрическими, так и непараметрическими методами. При анализе AD для QSAR моделей, в основном, используются непараметрические методы. Достоинством методов с применением функции плотности вероятности является возможность охарактеризовать пустые области внутри выпуклых оболочек.

При валидации моделей острой токсичности химических соединений с помощью AD используются следующие подходы: интервалы изменения переменных [28, 80, 106, 115, 117, 157, 178, 186, 196, 227, 232, 239, 311], проекции химических соединений на главные компоненты [91, 152, 211], Эвклидово расстояние [93, 98, 178, 311], леверидж [80, 93, 97, 161, 311], график Вильямса [97, 165], T^2 статистика Хотеллинга (Hotelling) [50], остаточное стандартное отклонение [48, 149], структурные фрагменты [144, 186, 196, 239], индекс надежности [172].

1.8. Модели

Как уже отмечалось выше, большинство исследований, связанных с построением QSAR моделей острой токсичности соединений, основано на использовании концепции MOA. При этом большое внимание уделяется разработке моделей наркоза [291]. Одной из причин повышенного интереса может служить тот факт, что большая группа используемых в промышленности веществ относится к этому типу MOA [6].

Многие органические соединения могут быть названы инертными [268], т.е. такими, которые не взаимодействуют специфически с рецепторами организма. Их токсичность может быть описана в виде простых зависимостей от $\log K_{ow}$ [289]. Такие соединения относятся к MOA неполярный наркоз. В качестве типичных представителей можно указать углеводороды, спирты, кетоны, эфиры. Каждое органическое соединение, в принципе, может действовать как наркотик, поэтому неполярный наркоз рассматривается в качестве базового или минимального токсического эффекта и соответствующие уравнения QSAR могут быть использованы для оценки минимальной токсичности химического соединения [290].

Другая группа соединений, таких как, фенолы, ароматические амины, являются менее инертными. Токсичность таких соединений также описывается простыми зависимостями от липофильности. Однако их величины токсичности

оказываются более высокими по сравнению с неполярными наркотиками при одних и тех же значениях $\log K_{ow}$. Такие химические соединения относят к группе полярных наркотиков. В литературе приводятся аргументы как в пользу разделения наркоза на два типа [291], так и против такого разделения [14, 292].

В пользу первой точки зрения свидетельствует факт существования двух разных прямых, описывающих токсичность неполярных и полярных наркотиков от $\log K_{ow}$. При исследовании синдрома острой токсичности у рыб (FATS) обнаружено, что соединения из обеих групп действуют обратимо, однако, у неполярных наркотиков эффект связан, в основном, с летаргией, в то время как у полярных наркотиков, с гиперактивностью. Изучение токсичности смесей соединений с МОА неполярный и полярный наркоз указывает на отсутствие аддитивности, что интерпретируется в пользу существования двух механизмов действия. Авторы работы [291] объясняют существование наркоза двух типов тем, что неполярные наркотики при взаимодействии с клеточными мембранами действуют по трем независимым направлениям (3D распределение), а полярные наркотики имеют 2D распределение.

Напротив, сторонники существования единого механизма наркоза указывают на то, что две наблюдаемые прямые в координатах $\log(LC_{50}) - \log K_{ow}$ для неполярных и полярных наркотиков можно свести к одной прямой при замене коэффициента распределения вещества в системе н-октанол – вода на коэффициент распределения мембрана-вода. Это, по их мнению, связано с неадекватностью используемого дескриптора (K_{ow}) для описания взаимодействия соединений с клеточными мембранами. Наличие единой корреляции между мембранными потенциалами и токсичностью для соединений, относящихся к МОА неполярный и полярный наркоз, также интерпретируется в терминах единого механизма наркотического действия.

Следует также отметить, что в литературе можно найти упоминания о других видах наркоза, например, аминном или эфирном наркозе [6]. При этом третичные амины хорошо соответствуют базовой модели. Доказательства существования эфирного наркоза выглядят спорными. В табл. 2 приведены

модели наркоза органических соединений по отношению к водным организмам с использованием одного дескриптора: $\log K_{ow}$.

Таблица 2

QSAR модели наркоза по отношению к водным организмам с использованием липофильности K_{ow}

Биообъект	Токсичность	Тип	Уравнение	n	r ²	s	F	q ²	Ссылка
C. decemmaculatus	LC ₅₀ , мг/л, 96 час.	NN	$\log(1/LC_{50}) = -2.08 + 0.375 \log K_{ow}$	9	0.804	-	-	0.705	86
C. sativus	EC ₅₀ , моль/л, 48 час.	PN	$\log(1/EC_{50}) = 1.99 + 0.92 \log K_{ow}$	29	0.850	0.15	144	-	221
C. vulgaris	EC ₅₀ , ммоль/л, 15 мин.	NN	$\log(1/EC_{50}) = -3.28 + 1.04 \log K_{ow}$	10	0.960	0.27	206	0.950	18
C. vulgaris	EC ₅₀ , ммоль/л, 15 мин.	PN	$\log(1/EC_{50}) = -1.91 + 0.641 \log K_{ow}$	10	0.880	0.16	69	0.840	18
D. magna	EC ₅₀ , моль/л, 48 час.	NN	$\log(EC_{50}) = -1.32 - 0.95 \log K_{ow}$	49	0.950	0.34	-	0.940	6
D. magna	LC ₅₀ , моль/л, 48 час.	NN	$\log(LC_{50}) = -1.28 - 0.86 \log K_{ow}$	36	0.940	0.44	311	0.900	6
D. magna	EC ₅₀ , ммоль/л, 48 час.	NN	$\log(EC_{50}) = 1.92 - 0.83 \log K_{ow}$	6	0.981	-	-	-	155
D. magna	LC ₅₀ , моль/л, 24 час.	NN	$\log(LC_{50}) = -1.093 - 0.773 \log K_{ow}$	17	0.912	0.24	-	-	232
D. magna	EC ₅₀ , моль/л, 48 час.	PN	$\log(EC_{50}) = -2.79 - 0.56 \log K_{ow}$	37	0.770	0.37	-	0.730	6
D. magna	LC ₅₀ , моль/л, 48 час.	PN	$\log(LC_{50}) = -2.21 - 0.80 \log K_{ow}$	33	0.740	0.45	90	0.940	6
D. magna	EC ₅₀ , мг/л, 48 час.	PN	$\log(EC_{50}) = -1.60 + 0.628 \log K_{ow}$	4	0.932	-	-	-	189
D. rerio	LC ₅₀ , мг/л, 96 час.	PN	$\log(LC_{50}) = 2.82 - 0.73 \log K_{ow}$	4	0.985	-	-	-	189
P. capricornutum	EC ₅₀ , моль/л, 96 час.	NN	$\log(EC_{50}) = -1.23 - 1.00 \log K_{ow}$	10	0.930	0.17	-	-	6
P. promelas	LC ₅₀ , ммоль/л, 96 час.	AN	$\log(1/LC_{50}) = -0.81 + 0.67 \log K_{ow}$	61	0.860	0.53	-	-	6

Продолжение табл. 2

Биообъект	Токсичность	Тип	Уравнение	n	r ²	s	F	q ²	Ссылка
<i>P. promelas</i>	LC ₅₀ , ммоль/л, 96 час.	EN	$\log(1/LC_{50}) = -0.64 + 0.64 \log K_{ow}$	14	0.950	0.22	207	-	6
<i>P. promelas</i>	LC ₅₀ , моль/л, 96 час.	GN	$\log(LC_{50}) = -1.74 - 0.81 \log K_{ow}$	144	0.880	0.45	-	0.870	6
<i>P. promelas</i>	LC ₅₀ , моль/л, 96 час.	NN	$\log(LC_{50}) = -1.39 - 0.85 \log K_{ow}$	58	0.940	0.36	-	0.930	6
<i>P. promelas</i>	LC ₅₀ , моль/л, 96 час.	PN	$\log(LC_{50}) = -2.16 - 0.72 \log K_{ow}$	86	0.900	0.33	-	0.900	6
<i>P. promelas</i>	LC ₅₀ , мг/л, 96 час.	NN	$\log(1/LC_{50}) = -2.49 + 0.457 \log K_{ow}$	9	0.812	-	-	0.691	86
<i>P. promelas</i>	LC ₅₀ , мкмоль/л, 96 час.	NN	$\log(1/LC_{50}) = -4.94 + 0.85 \log K_{ow}$	166	0.950	0.32	-	-	85
<i>P. promelas</i>	LC ₅₀ , мкмоль/л	NN	$\log(1/LC_{50}) = -4.25 + 0.94 \log K_{ow}$	60	0.940	-	-	-	85
<i>P. promelas</i>	LC ₅₀ , мкмоль/л	NN	$\log(1/LC_{50}) = -4.71 + 0.9 \log K_{ow}$	150	0.920	-	-	-	85
<i>P. promelas</i>	LC ₅₀ , мкмоль/л	NN	$\log(1/LC_{50}) = -3.93 + 0.63 \log K_{ow}$	10	0.860	-	-	-	85
<i>P. promelas</i>	LC ₅₀ , моль/л, 96 час.	NN	$\log(LC_{50}) = -1.107 - 1.027 \log K_{ow}$	16	0.808	0.55	-	-	232
<i>P. reticulata</i>	LC ₅₀ , моль/л, 96 час.	NN	$\log(LC_{50}) = -1.12 - 0.84 \log K_{ow}$	8	0.970	0.24	199	0.960	6
<i>P. reticulata</i>	LC ₅₀ , мг/л, 96 час.	NN	$\log(1/LC_{50}) = -2.51 + 0.477 \log K_{ow}$	9	0.858	-	-	0.775	86
<i>P. reticulata</i>	LC ₅₀ , моль/л, 96 час.	NN	$\log(LC_{50}) = -0.847 - 1.038 \log K_{ow}$	10	0.994	0.09	-	-	232
<i>P. reticulata</i>	LC ₅₀ , моль/л, 96 час.	PN	$\log(LC_{50}) = -2.00 - 0.76 \log K_{ow}$	11	0.890	0.28	72	0.840	6
<i>P. subcapitata</i>	EC ₅₀ , мг/л, 72 час.	PN	$\log(EC_{50}) = 2.04 - 0.479 \log K_{ow}$	4	0.369	-	-	-	189
<i>R. japonica</i>	LC ₅₀ , ммоль/л, 24 час.	PN	$\log(1/LC_{50}) = 1.35 + 0.95 \log K_{ow}$	21	0.910	0.21	195	-	201
River bacteria	EC ₅₀ , моль/л, 24 час.	NN	$\log(1/EC_{50}) = 2.000 + 0.565 \log K_{ow}$	10	0.887	0.07	62.6	-	215
<i>S. meliloti</i>	EC ₅₀ , ммоль/л	NN	$\log(1/EC_{50}) = -2.78 + 0.67 \log K_{ow}$	21	0.950	0.35	374	0.940	230
<i>T. pyriformis</i>	EC ₅₀ , ммоль/л, 40 час.	GN	$\log(1/EC_{50}) = -1.60 + 0.71 \log K_{ow}$	23	0.828	0.39	101	-	30

Продолжение табл. 2

Биообъект	Токсичность	Тип	Уравнение	n	r ²	s	F	q ²	Ссылка
<i>T. pygiformis</i>	EC ₅₀ , ммоль/л, 40 час.	NN	$\log(1/EC_{50}) = -2.02 + 0.78 \log K_{ow}$	87	0.960	0.20	2131	-	29
<i>T. pygiformis</i>	EC ₅₀ , ммоль/л, 40 час.	NN	$\log(1/EC_{50}) = -1.97 + 0.75 \log K_{ow}$	47	0.960	0.21	1132	-	239
<i>T. pygiformis</i>	EC ₅₀ , мг/л, 40 час.	NN	$\log(1/EC_{50}) = -1.79 + 0.723 \log K_{ow}$	215	0.926	0.27	-	-	28
<i>T. pygiformis</i>	EC ₅₀ , моль/л, 48 час.	NN	$\log(EC_{50}) = -1.067 - 0.737 \log K_{ow}$	28	0.901	0.21	-	-	232
<i>T. pygiformis</i>	EC ₅₀ , моль/л	PN	$\log(EC_{50}) = -1.979 - 0.629 \log K_{ow}$	153	0.780	0.38	520	-	72
<i>T. tubifex</i>	EC ₅₀ , моль/л, 3 мин.	NN	$\log(EC_{50}) = -0.179 - 0.848 \log K_{ow}$	30	0.955	0.16	-	-	232
<i>V. fischeri</i>	EC ₅₀ , ммоль/л, 15 мин.	NN	$\log(1/EC_{50}) = -3.580 + 1.307 \log K_{ow}$	12	0.933	0.20	140	-	132
<i>V. fischeri</i>	EC ₅₀ , ммоль/л, 15 мин.	NN	$\log(1/EC_{50}) = -1.14 + 0.818 \log K_{ow}$	75	0.712	0.73	184	-	306
<i>V. fischeri</i>	EC ₅₀ , ммоль/л, 15 мин.	PN	$\log(1/EC_{50}) = -1.359 + 0.935 \log K_{ow}$	22	0.879	0.18	146	-	132
Рыбы	LC ₅₀ , ммоль/л	NN	$\log(LC_{50}) = 1.29 - 0.691 \log K_{ow}$	7	0.975	-	-	-	155

Примечание.

AN - аминный наркоз, EN - эфирный наркоз, NN - неполярный наркоз, PN - полярный наркоз, GN = NN + PN

Кроме указанных в табл. 2 линейных регрессионных уравнений наркоза в зависимости от K_{ow} , существуют и другие регрессионные модели. В работе [6] представлена квадратичная модель неполярного наркоза, описывающая зависимость токсичности химических соединений от гидрофобности по отношению к *P. reticulata* (LC_{50} , мкмоль/л):

$$\log(1/LC_{50}) = -4.56 + 1.67 \log K_{ow} - 0.15 (\log K_{ow})^2$$

$n=11; r^2=0.890; s=0.27$

Аналогичное уравнение было использовано [170] для создания модели базовой (минимальной) токсичности спиртов и кетонов по отношению к грызунам (LD_{50} , ммоль/кг, перорально).

Крысы:

$$\log(1/LD_{50}) = -1.780(\pm 0.047) + 0.465(\pm 0.053) \log K_{ow} - 0.111(\pm 0.014) (\log K_{ow})^2$$

$n=55; r^2=0.59; s=0.15; F=40.3$

Мыши:

$$\log(1/LD_{50}) = -1.841(\pm 0.055) + 0.503(\pm 0.069) \log K_{ow} - 0.105(\pm 0.021) (\log K_{ow})^2$$

$n=30; r^2=0.72; s=0.17; F=38.5$

В качестве модели неполярного наркоза также используется билинейная модель. Данные по токсичности (крысы, LD_{50} , моль/кг, перорально) связаны с $\log K_{ow}$ следующим соотношением [240]:

$$\log(1/LD_{50}) = 1.13 + 0.6 \log K_{ow} - 0.8 \log(0.08 K_{ow} + 1)$$

$n=57; r^2=0.956; s=0.24$

Также известны модели наркоза, в которых наряду с $\log K_{ow}$ используются и другие дескрипторы. Например, уравнение [240], связывающее токсичность (крысы, LC_{50} , ppm, 4 час., ингаляция) соединений из различных химических классов (неполярный наркоз) с коэффициентами распределения вещества в системе н-октанол-вода и воздух-вода:

$$\log(1/LC_{50}) = -6.120 + 0.663 \log K_{ow} - 0.589 \log K_{aw}$$

$n=18; r^2=0.964; s=0.14$

В публикации [16] разработана модель неполярного наркоза (*Poesilia reticulata*, LC_{50} , мкмоль/л), в которую кроме $\log K_{ow}$ введен еще один дескриптор: DPSA-3 (разность площади поверхностей положительных и отрицательных зарядов третьего порядка):

$$\log(LC_{50}) = 5.43(\pm 0.10) - 0.95(\pm 0.02) \log K_{ow} - 0.039(\pm 0.006) \text{ DPSA-3}$$

$n=90; r^2=0.955; s=0.31; q^2=0.952$

Объектом исследования в работе [97] служили различные молекулярные дескрипторы при разработке новых статистически обоснованных QSAR моделей острой токсичности химических соединений по отношению к *P. promelas* (LC_{50} , моль/л, 96 час.). Всего было использовано около 1200 дескрипторов и с помощью GA были выбраны их лучшие комбинации. Валидация моделей проводилась как с использованием внутренних процедур (кросс-валидация, бутстреп, Y-рандомизация), так и с применением внешней процедуры (обучающая и тестовая выборки). В результате для группы соединений с МОА неполярный наркоз было получено уравнение:

$$\log(1/LC_{50}) = 2.6 + 0.72(\pm 0.02) \log K_{ow} - 0.13(\pm 0.02) E_{LUMO} - 1.03(\pm 0.2) \text{ RARS}$$

$n=263; r^2=0.95; s=0.28; q^2=0.95$

где RARS – GETAWAY дескриптор.

Полярный наркоз:

$$\log(1/LC_{50}) = 2.6 + 0.34(\pm 0.06) \log K_{ow} + 0.82(\pm 0.19) \text{BEHv3} + 0.18(\pm 0.04) \text{nHDon} - 0.65(\pm 0.19) \text{C-029}$$

$$n=86; r^2=0.90; s=0.36; q^2=0.88$$

где BEHv3 – самое большое собственное значение из матрицы Бурдена, nHDon – число Н-доноров, C-029 – центрированный атомный фрагмент. В работе особо подчеркивается необходимость использования для валидации полученных моделей химических соединений, которые не участвовали в построении этих моделей.

Важным фактором, который необходимо учитывать при QSAR моделировании токсичности, является ионизация соединений. Например, попытка построить общую регрессионную модель для 250 фенолов только на основе K_{ow} оказывается неудачной [208]. Статистически значимые уравнения удалось получить только путем группировки фенолов по заместителям и с использованием скорректированного (с учетом ионизации) значения коэффициента распределения вещества в системе н-октанол-вода. В публикации [64] для ряда резорцинолов, действующих как полярные наркотики на *T. pyriformis* (EC_{50} , мг/л, 40 час.), получено уравнение с хорошими статистическими характеристиками с использованием K_{ow} при pH=7.4:

$$\log(1/EC_{50}) = -1.27(\pm 0.09) + 0.83(\pm 0.04) \log K_{ow}$$

$$n=10; r^2=0.983; s=0.15; F=460; q^2=0.974$$

В работе [34] представлено регрессионное уравнение токсичности фенолов по отношению к *T. pyriformis* (полярный наркоз) с участием нескольких дескрипторов:

$$\log(\text{EC}_{50}) = -0.98(1.45) + 0.657(0.028) \log K_{ow} + 0.062(0.028) \text{pK}_a - 0.687(0.131) E_{\text{LUMO}} + 0.085(0.151) E_{\text{HOMO}} + 0.069(0.071) N_{\text{Hdon}}$$

$n=153; r=0.911; s=0.34; F=143$

Применение вместо MLR других методов статистического моделирования приводит к улучшению результатов: при использовании ANN $r=0.945$, а для SVM $r=0.947$. Модель для ряда фенолов и анилинов, действующих как полярные наркотики на *D. Magna* (EC_{50} , моль/л, 24 час.), представлена в публикации [143]:

$$\log(\text{EC}_{50}) = 5.68 + 0.437 \log K_{ow} - 0.0868 \text{pK}_a - 9.17 Q_H$$

$n=21; r^2=0.785; s=0.25; F=25.4$

где Q_H – максимальный заряд на атоме водорода.

Для другой группы фенолов и анилинов, токсичных по отношению к водоросли *S. obliquus* (EC_{50} , моль/л, 24 час.), получено уравнение с хорошими статистическими характеристиками [206]:

$$\log(1/\text{EC}_{50}) = 11.769 + 0.268 \log K_{ow} - 1.006 \Delta E$$

$n=20; r^2=0.946; s=0.17; F=148$

где $\Delta E = E_{\text{LUMO}} - E_{\text{HOMO}}$

В публикации [30] установлено, что замена K_{ow} на коэффициент распределения вещества в системе димиристоил фосфатидилхолин – вода (K_{DMPC}) приводит к улучшению общей регрессионной модели, предназначенной для описания токсичности химических соединений, относящихся к МОА неполярный и полярный наркоз (*T. pygiformis*, EC_{50} , ммоль/л, 40 час.):

$$\log(1/EC_{50}) = -1.62 + 0.73 \log K_{DMPC}$$

$$n=23; r^2=0.926; s=0.24; F=263$$

При моделировании токсичности соединений также находит применение линейная модель свободных энергий. Она использована для установления связи между токсичностью группы соединений с МОА неполярный и полярный наркоз и рядом сольватационных дескрипторов по отношению к трем простейшим: *E. sulcantum* (EC_{50} , моль/л, 72 час.), *U. parduczi* (EC_{50} , моль/л, 48 час.) и *C. paramecium* (EC_{50} , моль/л, 48 час.) [191] и пяти земноводным: *R. temporaria* (LC_{50} , моль/л), *R. pipiens* (LC_{50} , моль/л), *R. japonica* (LC_{50} , моль/л), *X. laevis* (LC_{50} , моль/л), *R. brevipoda porosa* (LC_{50} , моль/л) [216].

U. parduczi:

$$\log(1/EC_{50}) = 2.706(0.198) + 1.426(0.220) E + 0.433(0.212) S + 0.938(0.176) A - 1.025(0.337) B + 2.599(0.196) V$$

$$n=59; r^2=0.923; s=0.33; F=127$$

где E – избыточная молярная рефракция растворенного вещества, S – полярность/поляризуемость растворенного вещества, A – общая кислотность Н-связи растворенного вещества, B – общая основность Н-связи растворенного вещества, V – объем МакГовена (McGowan) растворенного вещества.

C. paramecium:

$$\log(1/EC_{50}) = 0.440(0.198) + 1.129(0.262) E + 0.160(0.232) S + 0.442(0.214) A - 1.826(0.315) B + 2.446(0.194) V$$

$$n=55; r^2=0.887; s=0.35; F=76.6$$

E. sulcantum:

$$\log(1/EC_{50}) = 0.489(0.203) + 0.894(0.072) E + 0.355(0.229) S + 1.108(0.222) A - 2.504(0.285) B + 2.852(0.182) V$$

$n=51$; $r^2=0.919$; $s=0.33$; $F=103$

R. temporaria:

$$\log(1/LC_{50}) = 0.629(0.085) + 0.821(0.083) E - 0.471(0.113) S - 0.082(0.108) A - 2.800(0.123) B + 3.165(0.094) V$$

$n=127$; $r^2=0.925$; $s=0.32$; $F=298$

R. temporaria+*R. pipiens*+*R. japonica*+*X. laevis*+*R. brevipoda porosa*:

$$\log(1/LC_{50}) = 0.716(0.075) + 0.703(0.082) E - 0.517(0.010) S + 0.127(0.086) A - 2.914(0.097) B + 3.308(0.075) V$$

$n=127$; $r^2=0.925$; $s=0.32$; $F=298$

Наряду с регрессионным анализом для построения QSAR моделей токсичности на основе МОА используются и другие методы. В работе [93] проведено PLS моделирование острой токсичности 311 органических соединений, действующих по механизму наркоза, на *P. promelas* (LC_{50} , ммоль/л, 96 час.). На основе использования пяти скрытых переменных были получены следующие результаты: для обучающего ряда соединений $n=208$, $r^2=0.909$, $s=0.343$ (логарифмические единицы) и для тестового ряда $n=103$, $r^2=0.888$, $s=0.424$.

Сравнение линейных и нелинейных методов моделирования токсичности для 250 структурно разнородных фенолов (5 МОА, *T. pyriformis*, EC_{50} , ммоль/л) проведено в работе [71]. Для описания структуры молекул было использовано 108 различных дескрипторов. Исходный ряд был разбит на обучающую (200 фенолов) и тестовую (50 фенолов) выборки. В качестве статистических методов исследования служили PLS и ANN (трехслойный перцептрон). По итогам работы делается вывод о преимуществе нелинейных методов моделирования перед линейными методами.

Для разработки количественных моделей острой токсичности химических соединений, ориентированных на МОА, необходима предварительная классификация соединений. Поэтому большое значение при QSAR моделировании играют классификационные методы. В работе [263] представлены результаты классификации 190 органических соединений, из которых 114 соединений относятся к МОА неполярный наркоз и 76 – к МОА полярный наркоз. С помощью нелинейного дискриминантного анализа в пространстве из 5 дескрипторов ($\log K_{ow}$, E_{LUMO} , E_{HOMO} , Q^+ , Q^-) проведена успешная классификация: степень правильного распознавания в первой группе составила $107/114=0.939$, а во второй группе $75/76=0.987$.

Модели классификации группы соединений, состоящей из наркотиков (48 соединений) и реактивных соединений (40) представлены в публикации [19]. С использованием 4-х дескрипторов ($\log K_{ow}$, E_{LUMO} , E_{HOMO} , S_{av}^N) исходные соединения были разделены на два класса при помощи двух методов: LDA и LR. При этом правильное распознавание в первой группе составило $42/48=0.875$ и $43/48=0.896$, во второй $36/40=0.900$ и $36/40=0.900$ для LDA и LR соответственно.

Наряду с относительно давно известными методами классификации используются и другие методы. Так, в работе [20] представлены результаты дискриминации рассмотренных выше 88 органических соединений на две группы: наркотики и реактивные соединения с применением SVM. Авторы отмечают важность учета двух моментов при использовании этого подхода: тип функции ядра и набор дескрипторов. Хорошие результаты классификации получены на основе полиномиальной функции второго порядка и 4-х дескрипторов ($\log K_{ow}$, E_{LUMO} , E_{HOMO} , S_{av}^N). В этом случае точность классификации составила 0.94 для обучающей выборки и 0.92 для тестовой выборки.

С использованием двух рядов органических соединений (T. pyriformis, 250 соединений, 5 МОА и P. promelas, 600 соединений, 6 МОА) была

разработана экспертная система для отнесения структур к одной из четырех групп в соответствии со схемой Верхаара [268], при этом исходное отнесение соединений было сделано вручную. Результаты классификации исходных рядов соединений свидетельствуют о плохой приспособленности схемы Верхаара для экспертной системы. Так, для обоих рядов соединений число молекул, правильно отнесенных к первой или второй группе, составляет всего лишь 38.8%.

В качестве примера успешного использования пошагового линейного дискриминантного анализа для классификации 221 фенола (Т. pyriformis, 4 MOA) можно привести данные работы [35]. Базовый набор состоял из следующих дескрипторов: $\log K_{ow}$, $\log D$, pK_a , E_{LUMO} , E_{HOMO} , число Н-доноров, число Н-акцепторов. В результате удалось сконструировать на основе 3-6 переменных LDA модели, у которых общее правильное распознавание составляло 86-89% для четырех механизмов токсического действия.

Исследованию классификации химических соединений с различными механизмами токсичности с использованием методов LDA, SVM, ANN посвящена публикация [44]. В работе было использовано два ряда соединений: 221 фенол (4 MOA) и 194 органических соединения (3 MOA). Установлено, что лучшие результаты классификации демонстрирует SVM с использованием радиальной базисной функции в качестве ядра. При использовании рассмотренного выше ряда фенолов для обучения нейронной сети со встречным распространением и последующей классификации предсказательная способность, оцененная с помощью кросс-валидации, составила 92% [50].

В работе [56] исследовано применение SVM для классификации 337 органических соединений (8 MOA). MOA классификация была основана на трех переменных: $\log K_{ow}$, $\log(1/EC_{50})$ (Т. pyriformis), $\log(1/LC_{50})$ (Р. promelas). Для получения классификационных моделей было исследовано большое число SVM моделей с различными ядрами. Установлено, что дискриминация между MOA неполярный наркоз и другими MOA может быть получена на основе

SVM модели с радиальной базисной функцией при уровне предсказательной точности 0.80.

Пример редко используемой для классификации процедуры логистической регрессии можно найти в публикации [65]. В качестве объекта классификации служила группа фенолов, состоящая из 153 полярных наркотиков (1 класс), 18 ингибиторов дыхания (2 класс), 27 проэлектрофилов (3 класс) и 23 мягких электрофилов (4 класс). Для дискриминации соединений было использовано пространство из четырех дескрипторов: $\log K_{ow}$, pK_a , E_{LUMO} , E_{HOMO} . Путем сравнения предсказанных MOA с предустановленными MOA был установлен процент неправильных предсказаний: 6.5% для 1 класса, 6.3% для 2 класса, 37.5% для 3 класса и 14.3% для 4 класса.

Рассмотренный выше ряд фенолов был использован для обучения и классификации с применением таких методов, как линейный дискриминантный анализ (LDA), различные регрессионные методы: линейная регрессия (LinReg), логистическая регрессия (LR), обобщенная аддитивная модель (GAM), сглаживание локально взвешенного регрессионного рассеивания (LOWESS), многомерные адаптивные регрессионные сплайны (MARS), проекционная направленная регрессия (PPR) [67]. В качестве переменных были использованы: $\log K_{ow}$, pK_a , E_{LUMO} , E_{HOMO} , N_{Hdon} (число доноров Н-связи). Общая ошибка предсказания составила 14.5%, 22.2%, 15.8%, 12.2%, 11.8%, 11.3%, 9.9% для LDA, LinReg, LR, GAM, LOWESS, MARS, PPR соответственно. На основе того же ряда фенолов (221 соединение, 4 MOA) была сконструирована экспертная система [68]. Степень правильных предсказания составила 92.2% для MOA полярный наркоз, 79.4% - ингибирование дыхания, 78.8% - проэлектрофильность, 71.1% - мягкая электрофильность.

Для рассмотренного выше ряда фенолов (4 MOA, 220 соединений (одно соединение было исключено)) был разработан двухэтапный пошаговый дискриминационный подход на основе методов LDA и LR [72]. При этом общее число используемых для дискриминации дескрипторов составляло 130.

Полученные результаты классификации оказались более высокими, чем у предшествующих моделей. Так, степень правильного распознавания для 1 класса составила 1.000 (0.961), для 2 класса – 0.700 (0.600), для 3-класса – 1.000 (0.846) и для 4 класса – 0.909 (0.909) при использовании LDA и LR соответственно.

Сравнение трех современных подходов для классификации: ILP (индуктивное логическое программирование), SVM (опорные вектора) и SVILP (индуктивное программирование на опорных векторах) проведено в работе [102]. Классификации подвергался ряд из 442 соединений (8 MOA, P. promelas). Установлено, что предлагаемый подход обладает более высокими классификационными характеристиками, чем традиционные подходы. Так, общая точность определения токсичности, рассчитанная на основе 5 кратной кросс-валидации составила 57.1%, 55.2% и 65.2% для методов ILP, SVM и SVILP соответственно.

В публикации [261] представлены результаты классификации 274 фенолов, относящихся к четырем типам токсического действия: полярные наркотики, ингибиторы дыхания, проэлектрофилы, мягкие электрофилы. Данные были разделены на обучающую (163 соединения) и тестовую выборки (111 соединений). Для описания структуры молекул было рассчитано 45 дескрипторов. В дальнейшем для работы было отобрано 9 дескрипторов. Для дискриминации химических соединений по типу действия было использовано четыре метода: SVM, KNN, ANN и AdaBoost (adaptive boosting). Было установлено, что алгоритм AdaBoost показывает более высокие результаты, чем другие методы, при классификации полярных наркотиков, проэлектрофилов и мягких электрофилов (общая степень распознавания составляла 94.1%, 100%, 77.8% соответственно). Однако в случае ингибиторов дыхания лучшие результаты демонстрировала нейронная сеть.

Метод дискриминантного анализа использован для классификации и предсказания трех типов наркоза: неполярный (143 соединения), полярный (44

соединений) и эфирный (7 соединений) наркоз [262]. В качестве дискриминирующих переменных были использованы R (избыточная молярная рефракция), π^H (диполярность/диполяризуемость), $\sum \alpha^H$ (суммарная кислотность Н-связи), $\sum \beta^H)^2$ (суммарная основность Н-связи). Ошибка классификации нелинейными дискриминантными функциями составила, по данным кросс-валидации, 2.1% для неполярного наркоза, 13.6% для полярного наркоза и 14.3% для эфирного наркоза.

В работе [264] выполнена дискриминация 115 химических соединений по 9 MOA. Структура молекул была описана с помощью 24-х дескрипторов. Классификация была проведена с использованием различных методов: LR, LDA, PLS, ANN. Выбор дескрипторов осуществлялся с помощью пошаговых методов или GA. Предсказательная способность моделей, оцененная на основе кросс-валидации, составила 52-59%. Для этого же ряда соединений был выполнен пошаговый дискриминантный анализ [275]. В пространстве из 10 физико-химических и структурных дескрипторов удалось правильно классифицировать 89.6% (103 из 115 соединений).

Следует отметить работу [90], в которой описана модель острой водной токсичности, для создания которой используется комбинация SVM и 18 переменных трех типов: дескрипторы молекулярного электростатического потенциала, топологические дескрипторы и $\log K_{ow}$. Модель включает в себя две схемы: классификация по уровню токсичности (TOXclass) и классификацию по MOA (MOAclass). В работе были использованы данные по острой водной токсичности (LC_{50} , 96 час., 9 MOA) по отношению к *P. promelas*. Тестовый ряд состоял из 296 соединений, а обучающий ряд - из 554, 263 соединений для TOXclass и MOAclass соответственно. Доля правильных предсказаний по данным кросс-валидации составила 84.3% для TOXclass и 77.2% для MOAclass.

Модели на основе MOA, в основном, используются для описания токсичности водных организмов. Это связано с тем, что QSAR модели, основанные на типах токсического действия, показывают более высокие

результаты, чем модели, основанные на химических классах [121]. Тем не менее, в литературе можно обнаружить работы, в которых исследуется токсичность с использованием химических классов. Например, в публикации [140] представлены регрессионные модели токсичности (*V. fischeri*, EC_{50} , ммоль/л, 30 мин.) для 138 галогензамещенных бензолов. Ряды соединений были сформированы с учетом типа галогена и места замещения. С использованием атомных полуэмпирических квантово-химических дескрипторов были получены регрессионные модели с хорошими статистическими показателями. Для земных организмов (мыши и крысы) основным подходом для формирования рядов соединений является использование химических классов.

Кроме регрессионных моделей токсичности, представленных в табл. 3, используются и другие подходы. Так, для ряда спиртов [7] в работе [168] предложена экспертная система для предсказания оральной токсичности 95 спиртов по отношению к мышам и крысам. Соединения были разбиты на четыре класса в зависимости от уровня токсичности. С использованием 27 правил было сконструировано дерево решений. На его основе была проведена классификация соединений. В результате степень правильного распознавания для обучающей выборки, состоящей из 95 спиртов, составила 0.91, а для тестовой выборки (25 спиртов) – 0.88.

Таблица 3

QSAR модели острой токсичности органических соединений по отношению к мышам и крысам

Объект	Токсичность	Соединения	Уравнение	n	r ²	s	F	q ²	Ссылка
К	LD ₅₀ , моль/кг, по.	Анилины	$\log(\text{LD}_{50}) = 2.55 - 2.27 \log K_{ow}$	8	0.865	0.05	36.5	-	7
К	LD ₅₀ , моль/кг, 24 час.	Асетамиды	$\log(1/\text{LD}_{50}) = 7.616 + 1.405 E_s^{\text{CH}} + 0.184(E_s^{\text{CH}})^2 + 0.600 I + 0.0186 \Delta H_f$	19	0.904	0.20	33.2	0.565	176
К	LD ₅₀ , ммоль/кг, по.	Гербициды	$\log(1/\text{LD}_{50}) = 16.76 + 1.29 \log K_{ow} - 0.30 (\log K_{ow})^2 - 4.13 \text{HD}$	12	0.884	0.17	-	-	7
К	LD ₅₀ , ммоль/кг, по.	Гербициды	$\log(1/\text{LD}_{50}) = -0.30 + 0.19 {}^2\chi^v + 0.36 s_1 - 0.07 \text{SHhother} - 0.70 \text{SdssC} - 0.98 \text{SHsOH}$	50	0.750	0.23	26.0	-	7
К	LD ₅₀ , моль/кг, по.	Гербициды	$\log(\text{LD}_{50}) = 0.16 S(R_2) + 0.08 \text{HBA}(R_2) - 0.26 \text{MR}(R_1)$	47	0.770	0.42	47.9	0.680	7
К	LD ₅₀ , ммоль/кг, по.	Произв. Бензола	$\log(1/\text{LD}_{50}) = -40.885 + 46.704 \text{DCW}$	14	0.915	0.37	130.0	-	177
К	LD ₅₀ , ммоль/кг, по.	Произв. Бензола	$\log(1/\text{LD}_{50}) = -10.2974 + 9.8114 \text{DCW}$	27	0.791	0.28	95.0	-	181
К	LD ₅₀ , ммоль/кг, по.	Разные	$\log(1/\text{LD}_{50}) = -1.41 + 0.00382 \text{TIE} - 0.645 \text{Hy} + 0.607 N_{\text{rings}6}$	26	0.815	0.53	32.4	0.735	227
К	LD ₅₀ , ммоль/м ² , вб.	Спирты	$\log(\text{LD}_{50}) = 3.51 - 0.75 {}^1\chi$	5	-	0.02	-	-	7
К	LD ₅₀ , ммоль/м ² , вв.	Спирты	$\log(\text{LD}_{50}) = 3.07 - 0.81 {}^1\chi$	8	-	0.14	-	-	7
К	LD ₅₀ , ммоль/кг, вб.	Спирты	$\log(\text{LD}_{50}) = 1.69 - 0.752 \log K_{ow} + 0.122 (\log K_{ow})^2$	8	0.973	-	-	-	169
К	LD ₅₀ , ммоль/кг, по.	Спирты	$\log(1/\text{LD}_{50}) = -1.780 + 0.465 \log K_{ow} - 0.111 (\log K_{ow})^2$	55	0.590	0.15	40.3	-	170

Продолжение табл. 3

Объект	Токсичность	Соединения	Уравнение	n	r ²	s	F	q ²	Ссылка
К	LD ₅₀ , моль/кг, по.	Спирты	$\log(1/LD_{50}) = 1.13 + 0.6 \log K_{ow} - 0.80 \log(0.08 K_{ow} + 1)$	57	0.956	0.24	-	-	240
К	LC ₅₀ , ppm, ин., 4 час.	Спирты	$\log(1/LC_{50}) = -6.120 + 0.663 \log K_{ow} - 0.589 \log K_{aw}$	18	0.964	0.14	-	-	240
М	LD ₅₀ , ммоль/кг, вв.	Алканы	$\log(LD_{50}) = 5.73 - 2.22 \log K_{ow} + 0.227 (\log K_{ow})^2$	4	0.867	-	-	-	169
М	LD ₅₀ , вб.	Алкилбензолы	$\log(1/LD_{50}) = -1.53 + 0.08 \log K_{ow}$	9	0.640	0.04	12.4	-	7
М	LD ₅₀ , вб.	Алкилбензолы	$\log(1/LD_{50}) = -1.37 + 0.16 \log K_{ow} - 0.04 (\log K_{ow})^2$	9	0.884	0.03	21.2	-	7
М	LD ₅₀	Барбитураты	$\log(1/LD_{50}) = -8.70 - 1.44 \log K_{ow} + 0.16 NVE$	11	0.924	0.08	-	0.879	237
М	LD ₅₀ , по.	Бенз. кислоты	$\log(LD_{50}) = 1.2399 {}^0J^A + 2.6911 {}^1J^A - 0.4445 {}^0J^A {}^1J^A$	39	0.986	-	-	-	164
М	LD ₅₀ , мг/кг, вб., 7 д.	Бензимидазолы	$\log(LD_{50}) = 7.380 - 5.719 \log K_{ow} + 1.404 (\log K_{ow})^2$	8	0.854	0.25	14.7	-	162
М	LD ₅₀ , ммоль/кг, по.	Кетоны	$\log(1/LD_{50}) = -1.841 + 0.503 \log K_{ow} - 0.105 (\log K_{ow})^2$	30	0.720	0.17	38.5	-	170
М	LD ₅₀ , вб.	Пиридины	$\log(1/LD_{50}) = 1.81 + 0.38 \log K_{ow} - 0.66 E_{LUMO}$	21	0.808	0.22	43.1	0.766	195
М	LD ₅₀ , ммоль/кг, по.	Разные	$\log(1/LD_{50}) = -1.12 + 0.00281 TIE - 0.790 Hy + 0.677 N_{rings6}$	26	0.797	0.57	28.8	0.751	227
М	LD ₅₀ , моль/кг, по.	Разные	$\log(1/LD_{50}) = 1.75 + 0.07 H-047 - 0.42 nR=Cs$	23	0.890	0.20	-	0.862	165
М	LD ₅₀ , ммоль/м ² , вб.	Спирты	$\log(LD_{50}) = 3.55 - 0.77 {}^1\chi$	5	-	0.07	-	-	7

Продолжение табл. 3

Объект	Токсичность	Соединения	Уравнение	n	r ²	s	F	q ²	Ссылка
М	LD ₅₀ , ммоль/м ² , вв.	Спирты	$\log(\text{LD}_{50}) = 3.16 - 0.80 {}^1\chi$	6	-	0.05	-	-	7
М	LD ₅₀ , ммоль/кг, по.	Спирты	$\log(\text{LD}_{50}) = 1.89 - 0.502 \log K_{ow} + 0.0994 (\log K_{ow})^2$	11	0.847	-	-	-	169
М+К	LD ₅₀ , ммоль/кг, вв.	Алкалоиды	$\log(1/\text{LD}_{50}) = 1.166 + 0.023 \text{MR}$	90	0.706	0.31	212.0	0.690	183
М+К	LC ₅₀ , ммоль/л, ин., 4 час.	Алканы	$\log(\text{LC}_{50}) = 3.05 - 1.80 \log K_{oa} + 0.232 (\log K_{oa})^2$	5	0.824	-	-	-	169
М+К	LC ₅₀ , ммоль/м ³ , ин., 4 час.	Нереактивные	$\log(\text{LC}_{50}) = 1.54 + 0.69 \log \text{VP}$	36	0.910	0.19	-	-	171
М+К	LC ₅₀ , ммоль/л, ин., 4 час.	Спирты	$\log(\text{LC}_{50}) = 3.13 - 1.14 \log K_{oa} + 0.0677 (\log K_{oa})^2$	5	0.982	-	-	-	169

Примечание.

⁰J^A - индекс связности 0 порядка, ¹χ – индекс молекулярной связности 1 порядка, ¹J^A - индекс связности 1 порядка, DCW - дескрипторы корреляционных весов, ΔH_f - теплота образования вещества, E^{CH}_s - скорректированный стерический параметр Тафта, Н-047 - атом-центрированный фрагмент, связанный с числом атомов водорода, НВА - протоноакцепторная способность, HD – жесткость (MNDO), Ну - гидрофильный фактор, I - индикаторная переменная, K_{aw} - коэффициент распределения вещества в системе воздух – вода, K_{oa} - коэффициент распределения вещества в системе н-октанол – воздух, MR - молекулярная рефракция, nR=C_s - число атомов углерода, имеющих sp² гибридизацию, N_{rings6} - число 6-членных колец, NVE - число валентных электронов, S – поверхность, TIE - топологический индекс, VP - давление пара, вб – внутрибрюшинно, вв – внутривенно, ин – ингаляционно, по – перорально, К – крысы, М – мыши.

Модели «количественная связь структура – токсичность» с использованием PLS и ANN были разработаны для оценки острой токсичности (LD_{50} , ммоль/кг) органофосфорных пестицидов по отношению к крысам [179]. Структура 51 соединения из обучающей выборки и 9 соединений из тестовой выборки была описана с помощью автокорреляционных векторов, кодирующих липофильность, молекулярную рефракцию, протонодонорную и протоноакцепторную способность молекул. На основе нейронных сетей были получены хорошие результаты: остаточное стандартное отклонение составило 0.29 и 0.26 для обучающей и тестовой выборок соответственно.

QSAR модели токсичности условно можно разбить на две группы. Самую большую группу составляют локальные модели, к которым можно причислить такие модели, в которых фигурирует десятки или сотни соединений и один МОА или химический класс. Ко второй группе можно отнести глобальные модели, описывающие сотни или тысячи соединений, в рамках нескольких МОА или химических классов. Для описания токсичности химических соединений сконструировано большое количество локальных QSAR моделей. Здесь, как, впрочем, и в других случаях, мы сталкиваемся с проблемой, связанной с числом используемых дескрипторов и возможностью их физико-химического истолкования. Интересным компромиссом, на наш взгляд, является подход, использующий поверхности отклика [221, 293]. В дальнейшем он был модифицирован [294] путем конструирования QSAR модели, основанной на двух дескрипторах, характеризующих гидрофобность ($\log K_{ow}$) и электрофильность (E_{LUMO}) соединений. Эта модель использована как для описания токсичности разнообразных фенолов, относимых к МОА полярный наркоз (*C. sativus*, EC_{50} , моль/л, 48 час.):

$$\log(1/EC_{50}) = 1.99 + 0.88 \log K_{ow} - 0.30 E_{LUMO}$$

$$n=29; r^2=0.93; s=0.10; F=173$$

так и для фенолов с различными типами действия (*C. sativus*, EC₅₀, моль/л, 48 час.) [223]:

$$\log(1/EC_{50}) = 2.40 + 0.65 \log K_{ow} - 0.72 E_{LUMO}$$

$$n=27; r^2=0.91$$

Также следует отметить попытку использования поверхности отклика для конструирования общей модели наркоза [62]. Для большого количества нейтральных наркотиков (*T. pyriformis*, EC₅₀, моль/л) была получена следующая модель:

$$\log(1/EC_{50}) = 1.50(\pm 0.037) + 0.68(\pm 0.013) \log K_{ow} - 0.13(\pm 0.013) E_{LUMO}$$

$$n=411; r^2=0.890; s=0.34; F=1652; q^2=0.888$$

Аналогичная модель была разработана для большой группы ароматических соединений с разными МОА (*T. pyriformis*, EC₅₀, 40 час.) [27]:

$$\log(1/EC_{50}) = -1.00 + 0.603 \log K_{ow} - 0.330 E_{LUMO}$$

$$n=239; r^2=0.800; s=0.34; F=476; q^2=0.796$$

Рассматриваемый ряд соединений был исследован также с помощью методов PLS и ANN. PLS модель была разработана на основе трех скрытых переменных и имела $r^2=0.790$ и $q^2=0.757$. ANN модель имеет более высокие статистические характеристики: $r^2=0.943$ и $s=0.06$. В качестве недостатков модели нейронных сетей отмечается отсутствие прозрачности и интерпретируемости.

В работе [28] уравнение поверхности отклика использовано для моделирования токсичности разнородных соединений с различными

механизмами действия (наркоз и др.) по отношению к *T. pygiformis* (EC₅₀, мг/л, 40 час.).

$$\log(1/EC_{50}) = -1.11(\pm 0.05) + 0.45(\pm 0.014) \log K_{ow} - 0.342(\pm 0.035) E_{LUMO}$$

$n=353; r^2=0.859; s=0.35$

Модель с учетом гидрофобности и электрофильности соединений [17] также разработана для группы соединений, относящихся как к неполярному наркозу, так и к специфическим типам токсического действия (*C. vulgaris*, EC₅₀, ммоль/л):

$$\log(1/EC_{50}) = -2.21(\pm 0.124) + 0.829(\pm 0.052) \log K_{ow} - 0.405(\pm 0.047) E_{LUMO}$$

$n=91; r^2=0.868; s=0.54; F=290; q^2=0.857$

Ингибирование роста дрожжей *S. cerevisiae* (EC₅₀, моль/л, 12 час.) для 24 нитроароматических соединений было исследовано в работе [204]. Все изученные соединения показывали избыточную токсичность по сравнению с токсичностью, соответствующей неполярному наркозу. Для описания связи «структура-токсичность» была использована модель:

$$\log(1/EC_{50}) = -0.46 + 0.41 \log K_{ow} - 0.89 E_{LUMO}$$

$n=24; r^2=0.87; s=0.22; F=72; q^2=0.86$

Иногда вместо дескриптора E_{LUMO} , характеризующего электрофильность соединений, используется A_{max} (акцепторная суперделокализация). В публикации [25] представлено уравнение, описывающее токсичность (*T. pygiformis*, EC₅₀, мг/л, 40 час.) большой группы соединений из различных классов и с разными типами токсического действия на основе гидрофобности ($\log K_{ow}$) и электрофильности (A_{max}):

$$\log(1/EC_{50}) = -5.91(\pm 0.20) + 0.545(\pm 0.015) \log K_{ow} + 16.21(\pm 0.62) A_{max}$$

$$n=385; r^2=0.859; s=0.27; F=1167; q^2=0.856$$

Аналогичная зависимость получена для ряда ароматических альдегидов (*T. pyriformis*) [57]:

$$\log(1/EC_{50}) = -4.04(\pm 0.85) + 0.583(\pm 0.034) \log K_{ow} + 9.80(\pm 2.62) A_{max}$$

$$n=52; r^2=0.864; s=0.20; F=156; q^2=0.844$$

Кроме локальных моделей токсичности в литературе имеются сведения о небольшом количестве глобальных моделей. В связи с этим, следует отметить публикацию [311], в которой представлены результаты совместной работы нескольких групп ученых по разработке моделей токсичности разнообразных химических соединений по отношению к *T. pyriformis*. В работе была использована одна и та же выборка из 983-х соединений, полученная в одной лаборатории, и которая случайным образом была разбита на обучающий ряд (644 соединения) и тестовый ряд (339 соединений). В качестве контрольной внешней выборки служил ряд из 110 добавочных соединений. Моделирование проводилось с применением различных подходов (KNN, SVM, MLR, PLS, ANN) и с использованием разных наборов дескрипторов. На основе обучающей выборки было разработано 15 различных QSAR моделей. Внутренняя валидация моделей проводилась с использованием LOO. При этом величина q^2 варьировала в пределах от 0.76 до 0.95. Коэффициент корреляции при внешней валидации r^2_{ext} изменялся от 0.71 до 0.85 и от 0.38 до 0.83 для тестовой и контрольной выборок соответственно. Учет AD при конструировании индивидуальных моделей ведет к улучшению характеристик валидации, но приводит к сокращению рассматриваемого химического пространства. Использование нового направления QSAR моделирования токсичности

химических соединений: консенсусных моделей приводит к улучшению предсказательных статистик и расширению химического пространства по сравнению с частными моделями.

К похожим выводам приходят авторы работы [178], в которой предложен комплексный подход для изучения острой токсичности (LD_{50}) 7385 разнообразных соединений при пероральном введении крысам. Используя 3472 соединения для обучения и 3913 соединения для валидации с помощью методов KNN, RF, HCA, FDA MDL QSAR и TOPKAT был получен ряд частных моделей токсичности. Сконструированная консенсусная модель при трех различных AD показывает лучшие результаты по сравнению с отдельными моделями. Так $r^2_{ext}=0.42$ и $r^2_{ext}=0.41$ для 74% соединений тестового ряда, $r^2_{ext}=0.48$ и $r^2_{ext}=0.41$ для 66% соединений и $r^2_{ext}=0.71$ и $r^2_{ext}=0.70$ для 19% соединений для консенсусной модели и лучшей частной модели соответственно.

Простая стратегия для предсказания токсичности соединений предложена в работе [180]. Она заключалась в том, что токсичность рассматриваемого соединения выражалась как среднее значение токсичностей для структурно родственных соединений. Структурное сходство устанавливалось на основе значений коэффициентов Танимото T_c . Авторами продемонстрирована возможность использования данного подхода на примере массива из 13645 разнообразных химических соединений, для которых известны величины LD_{50} (крысы, перорально). В качестве второго массива был использован ряд из 1781 соединений, идентифицированных как лекарства. При этом выявлена важность двух факторов, прямо влияющих на результаты предсказаний. Это пороговое значение T_c и минимальное число соседей. Предсказательная способность оценивалась с помощью кросс-валидации (LOO). Величина q^2 меняется при изменении величин T_c и числа соседей. Также при этом меняется число химических соединений, для которых может быть сделано предсказание. Например, для ряда из 13645 соединений при $T_c=0.75$ и одном соседе $q^2=0.82$,

число предсказанных токсичностей 57.3%. Для ряда из 1781 соединений при $T_c=0.75$ и одном соседе $q^2=0.74$, число предсказанных токсичностей 51.8%.

В работе [73] представлены результаты QSAR моделирования токсичности 250 фенолов (5 MOA) по отношению к *T. pyriformis* (EC₅₀, 40 час.) с использованием трех подходов: двухпараметровой поверхности отклика, пошаговой регрессии и искусственных нейронных сетей. Исходные данные были разбиты на обучающий ряд (200 соединений) и тестовый ряд (50 соединений). Для каждой молекулы было рассчитано 168 дескрипторов, отражающих физико-химические, структурные и топологические свойства. Модель токсичности с использованием поверхности отклика выглядит следующим образом:

$$\log(1/EC_{50}) = -0.24(\pm 0.075) + 0.42(\pm 0.029) \log D - 0.70(\pm 0.065) E_{LUMO}$$

$$n=250; r^2=0.54; s=0.56; q^2=0.53$$

Более высокими статистическими характеристиками обладает модель множественной линейной регрессии:

$$\log(1/EC_{50}) = 4.92(\pm 1.04) + 0.49(\pm 0.029) \log K_{ow} - 1.25(\pm 0.23) A_{Hard} + 0.22(\pm 0.058) NH_{Don} + 1.15(\pm 0.19) SdssC$$

$$n=250; r^2=0.66; s=0.48; q^2=0.64$$

где A_{Hard} – мера молекулярной жесткости, NH_{Don} – число доноров Н-связи, $SdssC$ – индекс электронного состояния атома углерода с двумя одиночными и двойной связью.

ANN моделирование проводилось на основе трехслойного перцептрона с использованием шести переменных: $\log K_{ow}$, $\log D$, Vol , ${}^0\chi^v_A$, ${}^4\chi_{PC}$, ${}^3\chi_P$, где Vol – молекулярный объем, ${}^0\chi^v_A$, ${}^4\chi_{PC}$, ${}^3\chi_P$, – индексы Кира и имело следующие статистические показатели: $r^2=0.71$ и $s=0.45$. Авторы отмечают, что нелинейные

методы имеют преимущество по сравнению с линейными методами. Однако ни один из используемых подходов не позволяет с приемлемой точностью предсказать токсичность соединений, действующих по механизму, включающего образование ковалентной связи.

Не привела к удовлетворительным результатам попытка моделирования острой токсичности 199 органических соединений, относящихся к 3 МОА: наркоз I, наркоз II и реактивные соединения (P. promelas, LC₅₀, мг/л, 96 час.) с помощью PLS анализа [95]. Данные были разделены на обучающую (159 соединений) и тестовую (40 соединений) выборки и на основе трех латентных переменных, полученных из 165 дескрипторов, была сконструирована модель, у которой $r^2=0.571$ и $s=0.76$ (обучающий ряд) и $r^2=0.246$ и $s=1.05$ (тестовый ряд).

В работе [97] представлена модель острой токсичности химических соединений (4 МОА, 249 соединений в обучающем ряду, 200 – в тестовом ряду) (P. promelas, LC₅₀, моль/л, 96 час.) на основе прямого предсказания токсичности. Этот подход включал в себя расчет около 1200 дескрипторов, отбор с помощью GA их лучших комбинаций и генерацию регрессионных уравнений. Полученная модель, по мнению авторов, является вполне удовлетворительной:

$$\log(1/LC_{50}) = 2.9(\pm 0.25) + 0.56(\pm 0.04) \log K_{ow} + 0.34(\pm 0.04) DP03 + 20.81(\pm 3.13) \\ H8m - 0.79(\pm 0.11) GATS1v - 1.59(\pm 0.35) R1v \\ n=448; r^2=0.81; s=0.57; q^2=0.80$$

где DP03 – молекулярный профиль Рандича, H8m, GATS1v, R1v – автокорреляционные дескрипторы.

Из-за ограниченного количества доступных экспериментальных токсикологических данных возникает необходимость дальнейшего развития методологии, связанной с их оценкой и экстраполяцией. В литературе достаточно широко представлен подход, основанный на исследовании

межвидовых корреляций токсичности [7, 18, 22, 52, 54, 55, 79, 123, 135, 152, 182, 186, 193, 220, 223, 278, 279]. В связи с этим, следует отметить публикацию [7], в которой выполнен обзор межвидовых корреляций острой токсичности химических соединений между мышами и крысами, с одной стороны, и различными биообъектами (такими как, бактерии, инфузории, рачки, рыбы и др.), с другой стороны. При этом в качестве моделей токсичности использованы как простые регрессионные зависимости, в которых в качестве зависимых и независимых переменных фигурируют только токсичности соединений, так и расширенные зависимости, в которых участвуют дополнительные дескрипторы.

Известны публикации, в которых представлены данные только для водных организмов. В случае близких организмов следует ожидать существование тесных корреляционных связей. Действительно, для ряда рыб (*O. mykiss*, *L. macrochirus*, *P. promelas*, *L. idus*, *Ictalurus sp.*, *Ciprinus sp.*) при использовании в качестве тестовых соединений большой группы пестицидов (свыше 600) обнаружены следующие соотношения для токсичности (LC_{50} , ммоль/л, 96 час.) [182]:

O. mykiss - *L. macrochirus*:

$$\log(LC_{50}) = -0.19 + 0.95 \log(LC_{50})$$

$$n=199; r^2=0.92; s=0.44; F=2168$$

O. mykiss - *L. idus*:

$$\log(LC_{50}) = -0.47 + 0.97 \log(LC_{50})$$

$$n=39; r^2=0.92; s=0.48; F=447$$

O. mykiss - *Ictalurus sp.*:

$$\log(LC_{50}) = -0.14 + 0.99 \log(LC_{50})$$

$$n=32; r^2=0.91; s=0.44; F=298$$

O. mykiss - *P. promelas*:

$$\log(LC_{50}) = -0.22 + 1.00 \log(LC_{50})$$

$$n=12; r^2=0.93; s=0.52; F=125$$

O. mykiss - *Ciprinus* sp.:

$$\log(LC_{50}) = -0.36 + 0.98 \log(LC_{50})$$

$$n=65; r^2=0.83; s=0.54; F=314$$

На основе токсикологических данных для *O. mykiss* и более далекого организма *D. magna* (EC_{50} , ммоль/л, 48 час.) была получена следующая модель:

O. mykiss – *D. magna*:

$$\log(LC_{50}) = -0.65 + 0.61 \log(EC_{50})$$

$$n=267; r^2=0.59; s=0.98; F=379$$

В случае похожей пары биообъектов *C. carpio* (LC_{50} , моль/л, 14 д.) – *D. magna* (EC_{50} , моль/л, 24 час.) и ограниченного ряда пестицидов было получено регрессионное уравнение более высокого качества [186]:

$$\log(LC_{50}) = -1.927 + 0.386 \log(EC_{50})$$

$$n=9; r^2=0.94; s=0.06; F=107$$

Для близких в биологическом отношении членистоногих *G. pulex* (LC_{50} , ммоль/л, 48 час.) и *D. magna* (EC_{50} , ммоль/л, 48 час.) характерно существование тесной корреляционной связи, как это следует из данных по токсичности для 44 химических соединений, обладающих 7 различными механизмами токсического действия [193]. Также установлено, что нет значительной разности между чувствительностью *G. pulex* и чувствительностью *D. magna* по отношению к той или иной МОА.

В работе [22] сконструирована простая модель межвидовой токсичности между *T. pyriformis* (EC₅₀, ммоль/л, 40 час.) и *P. reticulata* (LC₅₀, ммоль/л, 96 час.) с использованием соединений, относящихся к 3 МОА (нековалентный наркоз, нековалентная биореактивность, ковалентная биореактивность):

$$\log(1/LC_{50}) = 0.56 + 1.05 \log(1/EC_{50})$$

$$n=124; r^2=0.85; s=0.42; F=682$$

Аналогичная модель токсичности получена для другой пары организмов: *C. sativus* (RC₅₀, моль/л, 48 час.) и *T. pyriformis* (EC₅₀, моль/л) для группы соединений, состоящей из 16 полярных наркотиков и 5 реактивных соединений [223]:

$$\log(1/RC_{50}) = 1.63 + 0.93 \log(1/EC_{50})$$

$$n=21; r^2=0.87$$

Существование простой линейной зависимости между токсичностью 10 неполярных и 10 полярных наркотиков по отношению к инфузории *T. pyriformis* (EC₅₀, ммоль/л, 40 час.) и водоросли *C. vulgaris* (EC*₅₀, ммоль/л, 15 мин.) установлено в работе [18]:

$$\log(1/EC_{50}) = 0.61(\pm 0.10) + 0.81(\pm 0.05) \log(1/EC^*_{50})$$

$$n=20; r^2=0.93; s=0.30; F=244; q^2=0.92$$

Тесные корреляционные связи для ограниченного числа (14 индустриально значимых) органических соединений обнаружены между бактерией *V. fischeri* (EC₅₀, мг/л, 30 мин.) и рыбами (LC₅₀, мг/л, 96 час.), членистоногими *D. magna* (LC₅₀, мг/л, 48 час.), зелеными водорослями (LC₅₀,

мг/л, 95 час.) [135]. После удаления выбросов коэффициент корреляции (r^2) составил 0.95 для всех трех пар организмов.

В публикации [79] представлены данные по межвидовой корреляции токсичности 141 органического соединения для восьми водных организмов: бактерия *V. fischeri* (EC_{50} , моль/л, 15 мин.), дикие водные бактерии (EC_{50} , моль/л), водоросль *S. obliquue* (EC_{50} , моль/л, 48 час.), рачки *D. magna* (EC_{50} , моль/л, 24 час.), инфузория *T. pyriformis* (EC_{50} , моль/л, 40 час.), рыбы: *C. carpio* (LC_{50} , моль/л), *P. promelas* (LC_{50} , моль/л, 96 час.), *P. reticulata* (LC_{50} , моль/л, 96 час.). Результаты исследования показывают, что существуют хорошие межвидовые корреляции культурная бактерия/дикая бактерия ($r=0.92$) и рыба/рыба ($r=0.90\div 0.97$). Это предполагает, что изученные соединения обладают одинаковым механизмом действия для данных пар организмов. С другой стороны, были найдены плохие корреляции между токсичностью для пар *S. obliquue*/ *T. pyriformis* и *S. obliquue*/ *D. magna*, что указывает на разные механизмы токсического действия веществ на эти биообъекты. На основе полученных данных был также установлен порядок относительной чувствительности изученных организмов: *V. fischeri* $\approx$ *D. magna* $\approx$ рыбы (*C. carpio*, *P. promelas*, *P. reticulata*) $>$ водные бактерии $\approx$ *S. obliquue* $\geq$ *T. pyriformis*.

Большой объем работы выполнен авторами работы [279]. Ими проанализировано 5487 величин EC_{50}/LC_{50} для 180 биологических водных объектов и 1258 химических соединений. В результате было получено 780 корреляционных межвидовых моделей острой токсичности 550 соединений по отношению к 77 водным организмам. В работе было изучено влияние трех ключевых факторов на точность моделей: таксономической близости, МОА и параметров модели. В частности, было установлено, что модели, созданные на основе одного МОА, являются более устойчивыми, чем модели, построенные с использованием величин токсичности соединений из различных МОА. При этом также улучшается точность предсказания для пар биообъектов с большой таксономической дистанцией. Эти результаты показывают, что межвидовая

экстраполяция токсичности может быть улучшена с использованием МОА моделей для менее связанных таксономических пар.

Аналогичные выводы делаются в публикации [278], в которой проанализированы 4329 величин LD₅₀ при пероральном введении для 156 земных организмов (птицы, млекопитающиеся) и 951 химического соединения. При этом отмечается, что корреляции межвидовой токсичности являются средством для оценки чувствительности к загрязнению окружающей среды с известным уровнем точности для разнообразных организмов живой природы. В связи с этим контекстом разработаны корреляции токсичности между *D. magna* и рыбами чтобы оценить экотоксикологический потенциал опасности 77 разнообразных фармацевтических препаратов [152]. В работе представлены как отдельные, так и консенсусные модели, чтобы предсказать токсичность химических соединений по отношению к одному организму, когда доступны данные по другим организмам. Так, разработанные модели были использованы, чтобы предсказать токсичность для рыб 59 фармпрепаратов (для которых известны токсичности для *D. magna*) и токсичность для *D. magna* 30 фармпрепаратов (для которых известны токсичности для рыб).

В качестве общей модели для описания межвидовой токсичности наркотиков предложено уравнение поверхности отклика [54]. Для обучения были использованы данные по токсичности 125 соединений по отношению к *T. pyriformis* (EC₅₀, моль/л, 48 час.), 43-х соединений к *P. promelas* (LC₅₀, моль/л, 96 час.), 34-х соединений к *D. magna* (EC₅₀, моль/л, 24 час.). А для тестирования – 13, 12 и 9 наркотиков для *C. tarsalis* (LC₅₀, моль/л, 24 час.), *L. stagnalis* (LC₅₀, моль/л, 96 час.) и *R. temporaria* (минимальная наркотическая концентрация, моль/л) соответственно.

На основе обучающей выборки была получена общая модель, которая была использована для расчета токсичности соединений из тестового ряда:

$$\log (1/\text{Tox}) = b + 0.77(\pm 0.01) \log K_{ow} - 0.07(\pm 0.01) E_{LUMO}$$

$$n=202; r^2=0.96; F=1154; q^2=0.95$$

где $Tox=EC_{50}$, LC_{50} , $b=1.12(\pm 0.06)$, $1.49(\pm 0.04)$, $1.86(\pm 0.06)$, $0.93(\pm 0.15)$, $1.57(\pm 0.26)$, $1.02(\pm 0.04)$, $s^2=0.201$, 0.290 , 0.309 , 0.239 , 0.389 , 0.051 для *T. pyriformis*, *P. promelas*, *D. magna*, *C. tarsalis*, *L. stagnalis*, *R. temporaria* соответственно. Авторы работы отмечают, что точность данных, рассчитанных по общей модели, сопоставима с точностью, получаемой на основе уравнений для отдельных организмов. Полученная межвидовая модель для нековалентно действующих химических соединений легко интерпретируема и представляет собой удобный метод экстраполяции токсичности нековалентно действующих соединений от одного организма к другому.

Аналогичная общая модель была разработана на основе двух частных уравнений зависимости токсичности ряда альдегидов по отношению к *T. pyriformis* (EC_{50} , моль/л, 48 час.) и *T. promelas* (LC_{50} , моль/л, 96 час.) от физико-химических параметров [55]:

$$\log (1/Tox) = b + 0.505(\pm 0.033) \log K_{ow} + 14.315(\pm 3.731) D_{O-atom}$$

$$n=143; r^2=0.698; s^2=0.187; F=98.3; q^2=0.681$$

где D_{O-atom} – донорная делокализуемость на альдегидном атоме кислорода, $Tox=EC_{50}$, LC_{50} , $b=-1.812(\pm 1.117)$, $-1.172(\pm 1.113)$ для *T. pyriformis*, *P. promelas* соответственно. Полученная модель межвидовой токсичности была использована для аппроксимации токсичности альдегидов на другие биологические объекты: *P. reticulata*, *D. magna*, *V. fischeri*. При этом степень рассеивания экстраполированной токсичности сопоставима с рассеиванием, наблюдаемым в обучающей выборке.

1.9. Экспертные системы

Существует несколько экспертных систем, которые комбинируют многомерные QSAR модели или используют общие модели, чтобы предсказать токсичность химических соединений [6, 7, 9, 85, 90, 106, 112, 154, 167, 168, 196, 238, 295-297]. Так, TOPKAT [298] представляет из себя коллекцию классических SAR или QSAR моделей, полученных из дескрипторов, описывающих электронные, массовые и объемные свойства молекул. Для каждой модели определяется оптимальное предсказывающее пространство (OPS), вне пределов которого модель является неприменимой. Однако важно отметить, что наличие молекулы внутри OPS просто означает, что модель применима к молекуле, но это не гарантирует соответствия между предсказанным и экспериментальным значением активности [7, 9].

Система ECOSAR [299] основана на регрессионных моделях, использующих в качестве дескриптора $\log K_{ow}$. В настоящее время она насчитывает приблизительно 150 QSAR моделей для 50 различных структурных классов (например, нейтральные органические соединения, алифатические амины, сложные эфиры и т.д.) ECOSAR вообще хорошо работает для однородных рядов соединений, токсичность которых обусловлена общим нарушением клеточных мембран (наркоз I или неполярный наркоз) [106, 154].

Фундаментальным предположением MCASE [300] методологии является то, что наблюдаемая биологическая активность молекулы обусловлена наличием субструктур (биофоров). В ходе работы программа идентифицирует в обучающей выборке, состоящей из активных и неактивных молекул, статистически значимые субструктуры. Молекулы, содержащие эти «топ» биофоры, удаляются из базы данных. Процесс анализа повторяется для оставшихся молекул. Таким образом, MCASE расщепляет исходные данные на ряды, связанные с определенными биофорами. При этом предполагается, что

молекулы, входящие в состав этих рядов, действуют по одному механизму активности. Для каждого ряда MCASE идентифицирует добавочные параметры, называемые модуляторами, которые могут быть структурными или физико-химическими дескрипторами [7].

ASTER [301] является экспертной системой, в которой применяются химические структурные фрагменты чтобы предсказать вероятные типы токсического действия (MOA). Для MOA, которые имеют соответствующие QSAR модели, эта система оценивает острую токсичность по отношению к *P. promelas*. ASTER включает в себя модели, позволяющие оценить токсичность для неполярного, полярного и эфирного наркоза [106].

OASIS [302] использует модель поверхности отклика для предсказания острой токсичности для двух групп химических соединений: нековалентно (обратимо) действующих и ковалентно (необратимо) действующих. При этом применяются межвидовые корреляции острой токсичности по отношению к 17 водным организмам таким, как рыбы, улитки, головастики, гидроидные, ракообразные, личинки насекомых и бактерии [6].

Программа PASS [303] оценивает спектр биологического действия тестируемого соединения. В настоящее время возможно предсказание на основе структурной формулы соединения для 3678 видов биологической активности, включая токсические эффекты, со средней точностью предсказания около 95%. В основе используемого алгоритма лежит вероятностный подход Байеса.

В работе [106] проведено сравнение шести моделей для оценки острой токсичности 130 химических соединений по отношению к рыбе *P. promelas*. Для этого были использованы следующие компьютерные программы: ECOSAR, TOPKAT, ASTER, OASIS, нейронные сети PNN и CNN. Установлено, что наилучшей предсказательной способностью обладает PNN. TOPKAT также дает хороший прогноз токсичности в пределах оптимального предсказывающего пространства (OPS). Однако только 37% соединений из

тестового ряда попадает в OPS, что ограничивает применение этой программы при большом числе соединений.

В исследовании [153] предпринята попытка предсказать токсичность для 170 соединений из различных химических классов с использованием четырех моделей в качестве «черных ящиков». Оценка токсичности была сделана по отношению только к 122 соединениям, причем ECOSAR, TOPKAT и две отдельные QSAR модели для неполярного и полярного наркоза предсказали токсичность для 120, 39, 24 и 11 соединений соответственно. Авторы делают вывод о том, что имеющиеся QSAR модели токсичности не могут быть альтернативой использованию животных для получения экотоксикологических данных. Однако они могут использоваться в качестве средств, чтобы улучшить тестирование и уменьшить необходимое для тестирования количество животных.

С использованием компьютерной программы ECOSAR проведена оценка риска загрязнения окружающей среды фармпрепаратами путем предсказания их токсичности по отношению к рыбам, дафниям и водорослям [295, 296]. Установлен относительный порядок чувствительности организмов: дафнии > рыбы > водоросли.

С помощью MCASE авторами работ [85, 112] сконструированы модели острой токсичности широкого ряда из 219 органических соединений по отношению к рыбе *P. reticulata* и 685 соединений по отношению к рыбе *P. promelas*. Созданные модели обладают хорошей предсказательной способностью: они корректно оценивают острую токсичность для 80% органических соединений со средней ошибкой 0.63 и 0.40 логарифмических единиц от LC_{50} для *P. reticulata* и *P. promelas* соответственно. В работах продемонстрирована важность наркотического эффекта и установлены главные токсикофоры, соответствующие полярному наркозу и другим типам реактивных соединений.

Компьютерное моделирование токсичности выполнено с помощью системы PASS, QNA дескрипторов и самосогласованной регрессии [297]. В результате были разработаны QSAR модели острой токсичности LD₅₀ соединений по отношению к крысам для четырех типов введения вещества (перорально, внутривенно, внутрибрюшинно, подкожно). Используя тестовый ряд токсичности (LD₅₀, ммоль/кг, крысы, перорально) для 7286 соединений было проведено сравнение различных моделей. Установлено, что предлагаемый подход имеет более высокие статистические характеристики по сравнению с такими моделями, как иерархическая, ближайший сосед, FDA метод. Так, величина r^2_{test} составила 0.639, 0.573, 0.546, 0.555 соответственно.

В работе [196] представлена компьютерная программа KATE [304], предназначенная для оценки острой водной токсичности органических соединений. В качестве обучающей выборки используются экспериментальные данные LC₅₀ (96 час.) для *P. promelas* и *O. latipes* и EC₅₀ (48 час.) для *D. magna*. В системе KATE тестируемые соединения классифицируются в соответствии с их субструктурой, а уровень токсичности оценивается на основе соответствующей каждому классу модели токсичности. Они представляют собой простые регрессионные линейные зависимости между logK_{ow} и log(1/LC₅₀) или log(1/EC₅₀). Внутренняя кросс-валидация (LOO) этих моделей показывает, что уравнения, имеющие $r^2 > 0.7$, $s \leq 0.5$ и $n > 5$, обладают приемлемыми величинами q^2 . Внешняя валидация также демонстрирует низкую величину отклонений между экспериментальными и рассчитанными значениями токсичности.

Выводы по главе

Способность соединений к образованию водородной связи выражается на полуколичественном уровне в виде числа доноров и акцепторов, без учета молекулярного окружения. По-видимому, качество моделей токсичности, полученных с использованием таких дескрипторов, может быть значительно

улучшено при применении адекватных количественных характеристик протонодонорной и протоноакцепторной способности органических соединений.

Наиболее часто используемым дескриптором для конструирования моделей острой токсичности органических соединений, действующих по типу наркоз, является коэффициент распределения вещества в системе н-октанол-вода (K_{ow}). Однако, применение этого дескриптора приводит к появлению вопросов, связанных с объяснением различий между типами наркоза, главным образом, между неполярным и полярным наркозом. Одним из путей для решения этой проблемы может служить замена K_{ow} другими дескрипторами.

Число глобальных моделей токсичности, играющих важную роль при быстрой предварительной оценке токсичности большого числа разнообразных органических соединений, мало по сравнению с большим числом локальных моделей. Главным образом, это относится к изучению токсичности соединений по отношению к мышам и крысам.

Также можно сделать вывод о необходимости развития методологии QSAR. В первую очередь, это относится к разработке новых дескрипторов и способов установления связи между структурой и свойствами (активностью). Так, при описании структуры соединений, среди используемых дескрипторов полностью отсутствуют фрактальные дескрипторы, также не обнаружено работ, в которых для описания структуры и свойств (активности) используются двумерные кривые и применяются соответствующие подходы для установления связи между ними.

Глава 2. Экспериментальная часть

2.1. Физико-химические данные

В работе были использованы различные физико-химические данные, взятые из литературных источников и баз данных (табл. 4).

Таблица 4

Источники физико-химических данных

№	Свойство	Ед. изм.	Ссылка	Примечание
1	ΔH (энтальпия)	кДж/моль	[305]	T=298 K
2	ΔG (свободная энергия Гиббса)	кДж/моль	[305]	T=298 K
3	α (молекулярная поляризуемость)	$\text{\AA}^3$	[305]	-
4	K_{ow} (коэффициент распределения)	-	[306, 307]	-
5	d (плотность)	г/см ³	[308]	T=293 K
6	P (давление насыщенного пара)	мм Hg	[309]	-

2.2. Биологические данные

Данные по острой токсичности химических соединений были извлечены из различных литературных источников и баз данных. При этом исследовались биологические объекты с различной степенью сложности: начиная от бактерий и кончая грызунами.

В качестве характеристики острого токсического действия химических веществ по отношению к бактерии *Vibrio fischeri* использовались данные по снижению световой эмиссии EC_{50} , ммоль/л, со временем экспозиции $t = 5$ мин. [310]. При этом для 135 соединений было установлено 2 типа действия: неполярный наркоз и полярный наркоз. В случае инфузории *Tetrahymena pyriformis* токсический эффект оценивался на основе ингибирования их роста

EC₅₀, ммоль/л, t = 40 час. [239, 311], при этом 514 химических соединений были связаны с МОА неполярный наркоз [239].

Острая токсичность химических соединений по отношению к рыбам *Poecilia reticulata* [16], *Pimephales promelas* [238] и *Oncorhynchus mykiss* [312] выражалась в виде LC₅₀, ммоль/л, со временем экспозиции t = 96 час. В случае *Poecilia reticulata* для 293 соединений было установлено 4 МОА: неполярный наркоз, полярный наркоз, реактивный и особый [16]. Для 608 соединений и *Pimephales promelas* указано 8 МОА: наркоз I, наркоз II, наркоз III, ингибирование AChE, разобщение окислительного фосфорилирования, ингибирование дыхания, реактивный, захват центральной нервной системы [238]. Тип действия 131 соединения по отношению к *Oncorhynchus mykiss* был установлен нами по аналогии с имеющимися данными для других рыб и состоял из 3-х групп: неполярный наркоз, полярный наркоз, неопределенный тип.

В случае мышей и крыс вещества вводились внутривенно и мерой токсичности 10241 органических соединений служили величины LD₅₀, ммоль/кг, t = 24 час. [247]. В исходных данных МОА не указывались.

В качестве теста для оценки антигиббереллиновой (ретардантной) активности солей четвертичного аммония использовалось определение остаточной концентрации гибберелловой кислоты в клеточной культуре гриба *Gibberella fujikuroi* по методу [314] после инкубации в присутствии соответствующей соли в концентрациях $10^{-3} \div 10^{-6}$ моль/л.

2.3. Расчет, преобразование и отбор дескрипторов

2D дескрипторы химических соединений рассчитывались с использованием компьютерных программ HYBOT [305] (32 дескриптора), DRAGON [315] (3224 дескриптора) и DNESTR [316] (51 дескриптор). Расчет коэффициента распределения вещества в системе н-октанол-вода выполнялся

на основе компьютерной программы SLIPPER [317] (ClogP) или с помощью подхода Мorigучи [372] (MlogP). Стерические параметры E_s определены по методу, предложенному в работе [423]. В необходимых случаях проводилось автошкалирование дескрипторов [318] на основе формулы:

$$Z_{ij} = (X_{ij} - M_j)/s_j \quad (1)$$

где X_{ij} – исходное значение j -ого дескриптора для i -ой молекулы; M_j – среднее значение j -ого дескриптора; s_j – стандартное отклонение j -ого дескриптора; Z_{ij} – модифицированное значение j -ого дескриптора для i -ой молекулы.

Отбор дескрипторов для моделирования осуществлялся в два этапа. Сначала проводился анализ информативности переменных. Дескрипторы, которые имели свыше 95% постоянных значений, исключались из дальнейшего рассмотрения. Следующий этап селекции дескрипторов реализовывался с помощью подхода, основанного на использовании коэффициентов вариации и корреляции переменных. Коэффициент вариации для j -ого дескриптора вычисляли по модифицированной формуле [319]:

$$C_{vj} = |s_j/M_j| \quad (2)$$

Коэффициенты корреляции для дескрипторов j и k рассчитывали на основе соотношения [319]:

$$r_{jk} = \text{cov}_{jk}/(\text{cov}_{jj} \text{cov}_{kk})^{0.5} \quad (3)$$

где $\text{cov}_{jk} = (1/(n-1)) \sum (X_{ij} - M_j)(X_{ik} - M_k)$

Алгоритм отбора может быть описан в виде следующих шагов: 1) формирование матрицы данных (строки – молекулы, столбцы – дескрипторы); 2) отбрасывание малоинформативных дескрипторов, у которых свыше 95%

точек имеют постоянное значение; 3) расчет коэффициентов вариации; 4) определение дескриптора с максимальным значением C_v ; 5) расчет коэффициентов корреляции выбранного дескриптора с остальными дескрипторами; 6) формирование кластера вокруг этого дескриптора с включением переменных, коэффициент корреляции с которыми превышает заданное граничное значение; 7) расчет новой матрицы данных с отбрасыванием коррелированных переменных; 8) повторение шагов 4 – 7 до исчерпания списка дескрипторов.

Для расчета 3D дескрипторов использовались оптимизированные структуры, полученные с помощью программного комплекса HYPERCHEM [321]. В качестве стартовых молекул при проведении полной оптимизации служили генерируемые программой трехмерные структуры. Молекулы оптимизировали на основе полуэмпирического квантово-химического метода AM1 в приближении ограниченного метода Хартри - Фока с использованием алгоритма Полака - Рибера, среднеквадратичным градиентом 0.1 ккал/(Å моль), в вакууме.

Точечные группы симметрии молекул устанавливались с помощью информационной системы KNOWITALL [320] и HYPERCHEM [321].

2.4. Меры сходства химических соединений

В качестве мер структурного сходства химических соединений были использованы коэффициенты Танимото T_c , косинусные коэффициенты C_c и Эвклидовы коэффициенты E_c . Они рассчитывались с помощью компьютерной программы MOLDIVS [323] на основе структурных фрагментов с применением следующих формул:

$$T_c = N_{AB} / (N_A + N_B - N_{AB}) \quad (4)$$

$$C_c = N_{AB} / (N_A N_B)^{0.5} \quad (5)$$

$$E_c = 1 - ((N_A + N_B - 2 N_{AB}) / (2 N_{tot}))^{0.5} \quad (6)$$

где N_A – число структурных фрагментов в молекуле А, N_B – число структурных фрагментов в молекуле В, N_{AB} – число общих структурных фрагментов в молекулах А и В, N_{tot} – общее число структурных фрагментов в базе данных.

2.5. Статистические методы исследования

Для уменьшения размерности пространства и получения ортогонального набора дескрипторов использовалась компьютерная программа [324], реализующая метод анализа главных компонент (РСА). В качестве входной информации для анализа служили ковариационные или корреляционные матрицы. На выходе отбирались главные компоненты, вклад которых в общую дисперсию составлял не менее 1%.

В качестве основного метода классификации «без учителя» использовался иерархический кластерный анализ (НСА), реализованный в пакете статистических программ SPSS Statistics 17.0 [325]. Дистанционной мерой служило Эвклидово расстояние (для 4-х методов построения кластеров: межгрупповые связи, внутригрупповые связи, ближайший сосед, дальний сосед) или квадрат этого расстояния (для 3-х методов: центроидная кластеризация, медианная кластеризация, метод Варда). Число выделенных кластеров определяли путем анализа зависимости коэффициента Е от числа кластеров на каждом шаге агломерации с использованием стандартной методики SPSS, основанной на выявлении резких изломов на анализируемой кривой. Коэффициент Е представлял из себя расстояние между парой объединяемых кластеров. Способ оценки этого расстояния зависил от метода построения кластеров. Так, в случае межгрупповых связей, это расстояние равно среднему значению всех расстояний между всеми возможными парами точек из обоих кластеров.

Для проведения классификации «с учителем» применялись различные методы, включая линейный дискриминантный анализ (LDA), метод k -ого ближайшего соседа (KNN), метод структурного сходства (SS), метод опорных векторов (SVM), искусственную нейронную сеть (ANN).

Линейный дискриминантный анализ осуществлялся на основе библиотеки компьютерных программ IMSL [326]. LDA проводился с использованием объединенной ковариационной матрицы, единичных частот и весов, пропорциональных априорных вероятностей, в качестве метода классификации использовалась реклассификация.

Метод k -ого ближайшего соседа также выполнялся с помощью процедур IMSL [326]. Дискриминация с применением метода KNN была выполнена с использованием евклидовой метрики, пропорциональных априорных вероятностей и порога апостериорных вероятностей равным 0.1. Максимальный номер ближайшего соседа устанавливался равным 20.

Классификация с применением метода SS проводилась на основе собственной компьютерной программы, реализующей подход, описанный в работе [314]. В качестве меры структурного сходства использовался коэффициент Танимото. Максимальное число рассматриваемых соседей составляло 5.

Для выполнения дискриминации с помощью метода опорных векторов применялась программа LIBSVM [328]. При этом использовались стандартный алгоритм классификации C-SVC и радиальная базисная функция в качестве ядра. Оптимальные значения параметров γ и C определялись с применением кросс валидации обучающей выборки и решеточного метода.

На основе SPSS [325] в качестве средства классификации была использована искусственная нейронная сеть. При этом ANN представляла из себя многослойный перцептрон с автоматическим выбором параметров архитектуры. Количество скрытых слоев было равно 1, в качестве функции активации выступал гиперболический тангенс. В выходном слое такой

функцией служил softmax (функция активации, специально предназначенная для классификационных сетей с кодированием по методу один-из-N).

Множественный регрессионный анализ проводился с применением модифицированной нами компьютерной программы SVD [329]. Модификация заключалась в добавлении процедур расчета статистических характеристик и проведения перекрестного контроля с выбором по одному. Для анализа коллинеарности независимых переменных использовалось два подхода. Первый из них применялся до проведения регрессионного анализа и заключался в вычислении парных коэффициентов корреляции. Если величина r^2 превышала критическое значение (обычно 0.5), одна из переменных отбрасывалась. Другой подход использовался после проведения процедуры MLR и сводился к анализу сингулярных чисел независимых переменных.

2.6. Валидация моделей QSAR

В качестве базовых характеристик качества бинарной классификации служили величины TP, FN, TN и FP. TP – это число верных обнаружений в первой группе, FN – число ложных обнаружений в первой группе, TN – число верных обнаружений во второй группе, FP – число ложных обнаружений во второй группе. На основе этих характеристик были рассчитаны величины, характеризующие описательную способность моделей классификации [330]:

чувствительность:

$$SN = TP / (TP + FN) \quad (7)$$

специфичность:

$$SP = TN / (TN + FP) \quad (8)$$

точность:

$$ACC = (TP + TN)/(TP + FN + TN + FP) \quad (9)$$

коэффициент корреляции Мэтьюза [331]:

$$MCC = (TP \cdot TN - FP \cdot FN) / ((TP + FP)(TP + FN)(TN + FP)(TN + FN))^{0.5} \quad (10)$$

Описательная способность регрессионных моделей оценивалась с помощью следующих величин: размера выборки (n), коэффициента линейной корреляции (r), остаточного стандартного отклонения (s) и критерия Фишера (F). Для их расчета были использованы формулы [332]:

$$r^2 = 1 - PRESS/TSS \quad (11)$$

где $PRESS = \sum(Y_{\text{эксп}} - Y_{\text{расч}})^2$; $TSS = \sum(Y_{\text{эксп}} - Y_{\text{сред}})^2$; $Y_{\text{сред}} = \sum Y_{\text{эксп}}/n$, здесь $Y_{\text{эксп}}$ – экспериментальные значения зависимой переменной, $Y_{\text{расч}}$ – расчетные значения зависимой переменной, $Y_{\text{сред}}$ – среднее значение зависимой переменной.

$$s = (PRESS/(n - m))^{0.5} \quad (12)$$

где m – число коэффициентов в регрессионной модели.

$$F = r^2 \cdot (n - m) / ((1 - r^2) \cdot (m - 1)) \quad (13)$$

Для определения прогностической способности моделей QSAR использовалась внутренняя и внешняя валидация. Внутренняя валидация проводилась с использованием перекрестного контроля с выбором по одному и количественной меры в виде коэффициента линейной корреляции q^2 [333]. Формула для расчета этой величины выглядит следующим образом:

$$q^2 = 1 - PRESS^*/TSS^* \quad (14)$$

где $PRESS^* = \sum (Y_{\text{эксп(обуч)}} - Y_{\text{расч(обуч)}})^2$; $TSS^* = \sum (Y_{\text{эксп(обуч)}} - Y_{\text{сред(обуч)}})^2$; $Y_{\text{сред(обуч)}} = \sum Y_{\text{эксп(обуч)}}/n$, здесь $Y_{\text{эксп(обуч)}}$ – экспериментальные значения зависимой переменной в обучающей выборке, $Y_{\text{расч(обуч)}}$ – расчетные значения зависимой переменной в обучающей выборке, $Y_{\text{сред(обуч)}}$ – среднее значение зависимой переменной в обучающей выборке.

Для проведения внешней валидации использовался метод, который заключался в разделении анализируемых данных на обучающую и тестовую выборки в пропорции 80% и 20% соответственно. Алгоритм формирования выборок с использованием программы MOLDIVS [323] состоял в следующем: 1) генерирование библиотеки структурных фрагментов; 2) расчет мер структурного различия (dissimilarity) молекул на основе T_c ; 3) расчет величин структурного разнообразия (diversity) молекул на основе мер структурного различия; 4) упорядочивание молекул по величине структурного разнообразия; 5) выбор каждой пятой молекулы для создания тестовой выборки. Количественной оценкой внешней валидации служил следующий коэффициент корреляции [335]:

$$r_{\text{пред}}^2 = 1 - PRESS^{**}/TSS^{**} \quad (15)$$

где $PRESS^{**} = \sum (Y_{\text{эксп(тест)}} - Y_{\text{расч(тест)}})^2$; $TSS^{**} = \sum (Y_{\text{эксп(тест)}} - Y_{\text{сред(обуч)}})^2$; $Y_{\text{сред(обуч)}} = \sum Y_{\text{эксп(обуч)}}/n$, здесь $Y_{\text{эксп(тест)}}$ – экспериментальные значения зависимой переменной в тестовой выборке, $Y_{\text{расч(тест)}}$ – расчетные значения зависимой переменной в тестовой выборке, $Y_{\text{сред(обуч)}}$ – среднее значение зависимой переменной в обучающей выборке, $Y_{\text{эксп(обуч)}}$ – экспериментальные значения зависимой переменной в обучающей выборке.

Область применимости моделей определялась двумя способами. Один из них состоял в простом вычислении минимальных и максимальных значений дескрипторов и токсичности в обучающей и тестовой выборке. При

использовании описанного выше метода формирования выборок на основе структурного разнообразия наблюдалось практически полное перекрывание интервалов изменения переменных в обеих выборках.

Второй подход для определения АД был реализован на базе компьютерной программы AMBIT [334]. Расчеты выполнялись с использованием метода плотности вероятности с граничными значениями вероятности 95 и 75%.

2.7. Методика проведения спектральных и калориметрических измерений

ИК-спектры твердых и жидких соединений регистрировали на приборе Specord 75 IR в области $600\text{--}4000\text{ см}^{-1}$. Точность определения максимумов полос поглощения $\pm 5\text{ см}^{-1}$. Спектры твердых образцов снимались в виде суспензии в вазелиновом масле, которая помещалась между пластинами из KBr. Спектры жидких образцов снимали при помещении капли вещества между пластинами из KBr или ими наполняли кюветы из CaF_2 .

ИК спектрофотометрические измерения растворов соединений проводили на приборе Perkin-Elmer 283B в области валентных колебаний групп OH, NH в режиме определения оптической плотности. Точность определения максимумов полос поглощения $\pm 2\text{ см}^{-1}$. При этом использовались кюветы из CaF_2 толщиной 0.50-2.87 мм. В кювету сравнения помещался чистый растворитель, в некоторых случаях раствор протоноакцепторного соединения.

Калориметрические измерения были проведены на дифференциальных микрокалориметрах с изотермической оболочкой LKB 2107/112 и ДАК-1-1А с погрешностью измерений $1 \div 2\%$.

Теплоты реакций изучались с использованием методики смешения растворов реагентов или калориметрического титрования. Калориметр имеет два сосуда: один рабочий, другой сравнения, каждый из которых имеет два отделения. После термостатирования при 298 К растворы помещались в

калориметр. При использовании первого метода в отделения рабочего сосуда заливались исследуемые растворы веществ, а в отделения сосуда сравнения – растворитель и один из рабочих растворов. Таким образом, происходил автоматический учет теплоты разбавления одного из реагентов. Теплота разбавления другого реагента определялась в дополнительном опыте. При использовании второго метода в рабочий сосуд помещался раствор одного из реагентов (титруемое вещество), в сосуд сравнения – растворитель. Раствор другого реагента (титранта) заливался в микрошприцы, соединенные с обоими сосудами. После термостатирования приблизительно в течение часа растворы смешивались. При использовании методики калориметрического титрования непосредственно перед смешением определенное количество титранта добавлялось в сосуды.

Калориметрический сигнал пропорциональный разности тепловыделений в сосудах регистрировался самописцем, интегральный сигнал которого $Z_{\text{эксп}}$ пропорционален теплоте $Q_{\text{эксп}}$. Коэффициент пропорциональности K определялся с помощью встроенного внутрь реакционного сосуда джоуль-калибратора с известным сопротивлением R , пропуская через него известный ток I в течение времени t , регистрируя интегральный сигнал калибровки Z_k . Теплота опыта рассчитывалась по формуле:

$$Q_{\text{эксп}} = K Z_{\text{эксп}} = I^2 R t Z_{\text{эксп}}/Z_k \quad (16)$$

2.8. Планирование эксперимента и оценка термодинамических параметров водородной связи по данным экспериментальных измерений

Для корректной оценки термодинамических параметров при Н-связывании экспериментальную работу необходимо проводить в определенной области изменения концентрации реагентов. В своем исследовании мы осуществляли планирование эксперимента используя подход, предложенный в

публикации [336]. В этой работе показано, что константы образования комплексов состава 1:1 и специфические для каждого из используемых методов исследования интенсивные параметры (ИК спектроскопия – молярный коэффициент поглощения, ЯМР спектроскопия – химический сдвиг и т.д.) определяются с минимальными ошибками, если, так называемая, «степень насыщения» (S), определяемая по формулам (17)-(18) меняется в пределах $0.2 \div 0.8$.

Для компонента А:

$$S_a = C_{ab}/C_a \quad (17)$$

для компонента В:

$$S_b = C_{ab}/C_b \quad (18)$$

где C_{ab} – концентрация комплекса АВ, C_a – общая концентрация компонента А, C_b – общая концентрация компонента В. С помощью ИК спектроскопии определялась область концентраций протонодонора, при которой его самоассоциацией можно было пренебречь. Затем по 2-3 точкам спектрофотометрически по известному уравнению (19):

$$K = (1 - D^*/D_0)/((D^*/D_0) (C_b^0 - C_a^0 (1 - (D^*/D_0)))) \quad (19)$$

где D_0 и D^* – оптические плотности для полосы поглощения $\nu_{(AH)}$ свободного протонодонора до и после комплексообразования, оценивалась константа Н-связывания (K). После этого по программе, реализующей алгоритм [336], для данной концентрации одного из реагентов (титруемое вещество) определялась необходимая область изменения концентрации другого реагента (титрант). На

основе полученной информации планировался калориметрический эксперимент.

Расчет стехиометрии комплексов, находящихся в равновесии с реагентами, логарифмов констант равновесий и энтальпий комплексообразования проводился с использованием двух алгоритмов минимизации функционала (20) исходя из разных начальных приближений:

$$F = \sum (Y_{\text{эксп}} - Y_{\text{расч}})^2 \quad (20)$$

где $Y_{\text{эксп}}$ – экспериментальное значение свойства равновесной системы или характеристики процесса, $Y_{\text{расч}}$ – расчетное значение свойства равновесной системы или характеристики процесса. Первый алгоритм EQ [337], использующий модифицированный метод Ньютона, был расширен применительно к калориметрическому титрованию. Вторым алгоритмом SEARCH основан на симплекс-методе Нелдера и Мида [338]. Для оценки адекватности полученных решений применяли статистические критерии: стандартное отклонение, R – фактор [339] и проводили анализ остатков [332].

2.9. Исходные вещества, реактивы и их квалификация

Растворители: гексан «ч», бензол «хч» и четыреххлористый углерод «чда» очищали по стандартным методикам [340]. Фенол «чда» использовали без дополнительной очистки. Гексаметилфосфортриамид «ч» выдерживали несколько суток над СаО, затем над молекулярными ситами 4А, перегоняли в вакууме над СаО и СаН₂. Контроль чистоты соединений осуществляли методами газовой хроматографии и ИК спектроскопии. Синтез и очистка фосфорилированных оксимов и дифторнитроацетанилидов описаны в работах [341] и [342] соответственно.

Глава 3. Количественные модели «структура – физико-химические свойства» органических соединений

3.1. Аддитивно-мультипликативная модель энтальпии и свободной энергии водородной связи. Монофункциональные соединения

Взаимодействие между протонодонорными и протоноакцепторными молекулами приводит к образованию различных молекулярных и ионных Н-комплексов, которые играют важную роль в химических процессах, включая ферментный катализ [343]. Более того, значение водородной связи особенно проявляется в биологических системах, где она выступает в качестве критического структурного элемента или определяющей силы в молекулярном распознавании [344-346]. Как уже указывалось ранее, в литературе также имеются сведения об участии Н-связи при формировании эффекта токсичности. Однако методы оценки протонодонорной и протоноакцепторной способности молекул оставляют желать лучшего: в большинстве случаев в качестве такой меры выступает просто число Н-доноров и Н-акцепторов.

Надежный путь количественного описания водородной связи заключается в использовании термодинамических функций образования Н-комплексов. Следовательно, этот подход может быть полезным и при QSAR моделировании. Энтальпия (ΔH), свободная энергия Гиббса (ΔG) и энтропия (ΔS) образования Н-комплексов связаны между собой хорошо известным соотношением:

$$\Delta G = -RT \ln K = \Delta H - T \Delta S \quad (21)$$

Можно выделить два главных подхода для оценки термодинамических функций водородной связи: 1) квантово-химические расчеты, моделирование Монте-Карло или молекулярная динамика и 2) корреляционный анализ. Второй подход основан на использовании различных наборов параметров (в частности,

шкалы Н-связи) и специальных аналитических уравнений. Разработка и развитие эмпирических расчетов термодинамических функций Н-связи начались в середине шестидесятых годов двадцатого века. В это время было предложено два подхода: аддитивно-мультипликативный и мультипликативный.

Аддитивно-мультипликативная схема расчета энтальпии донорно-акцепторного взаимодействия предложена Драго и сотр. [347-349]:

$$-\Delta H = e_A e_B + c_A c_B + t_A t_B \quad (22)$$

где параметры e указывают на тенденцию льюисовской кислоты (А) или основания (В) подвергаться электростатическому взаимодействию, c – ковалентному и t – взаимодействию с переносом заряда. Указанная схема дает неплохие оценки величин ΔH и поэтому получила достаточно широкое распространение. В качестве основного недостатка формулы Драго следует рассматривать невозможность сопоставления относительной электронодонорной и акцепторной способности атомов и групп вследствие того, что величина энтальпии оценивается через несколько параметров.

В этом отношении полезными оказываются еще более простые формулы, основанные на предположении о постоянстве донорных и акцепторных функций взаимодействующих центров. Здесь, прежде всего, необходимо указать на работы Иогансена [350-353], в которых развито количественное представление о так называемом правиле факторов, и аналогичные работы Терентьева [354]. Правило факторов для энтальпий и интенсивностей полос в ИК спектрах Н-комплексов было предложено первоначально в работе [350] и получило развитие в работах [351-353]. Согласно правилу факторов энтальпия водородной связи пропорциональна произведению двух чисел: фактора i кислоты (P_i) и фактора j основания (E_j):

$$\Delta H_{ij} = \Delta H_{11} P_i E_j = -22.2 P_i E_j \quad (23)$$

где коэффициент пропорциональности ΔH_{11} – энтальпия произвольного стандартного Н-комплекса, для которого положено $P_i=E_j=1$. Иогансен выбрал в качестве стандартов фенол и диэтиловый эфир, для которых в среде CCl_4 при 298 К по данным указанного автора $-\Delta H_{11}=22.2$ кДж/моль. Безразмерные P_i и E_j факторы (дескрипторы) характеризуют относительную энергетическую способность данной кислоты или основания к образованию Н-связей с любым основанием или кислотой. Факторы взаимно независимы и постоянны. При таком подходе $M+N$ параметров равных суммарному числу кислот (M) и оснований (N) позволяют предсказать энергии $M \times N$ различных Н-комплексов.

Развитию подхода Иогансена посвящены работы [355, 356]. Для различия электронодонорных и электроноакцепторных функций соединений их факторам были приписаны разные знаки (для электронодонорного знак плюс, для электроноакцепторного – минус). Это дало возможность переписать формулу (23) в следующем виде:

$$\Delta H_{ij} = |\Delta H_{11}| E_i E_j \quad (24)$$

где $E_i = -P_i$. Такая запись выглядит вполне естественно с позиции физической химии, поскольку при Н-связывании кислот и оснований произведение факторов с противоположными знаками приводит к экспериментально наблюдаемым отрицательным значениям ΔH . В дальнейшем для некоторых классов сильных Н-доноров и Н-акцепторов нами было предложено использовать формулу (24) в модифицированном виде:

$$\Delta H_{ij} = |\Delta H_{11}| E_i E_j + k_0 \quad (25)$$

Наши последующие исследования показали, что применение только энтальпии в качестве параметра недостаточно для адекватного описания

способности соединений к образованию Н-связи. Поэтому мы попытались использовать правило факторов и по отношению к энтропии образования Н-комплексов. Результаты оказались вполне удовлетворительными. Однако, учитывая, что непосредственно из эксперимента определяются константы равновесия, мы применили мультипликативный подход непосредственно к ΔG . Для этого из литературы были извлечены данные по константам и энтальпиям образования 174 Н-комплексов состава 1:1 в CCl_4 для различных соединений. Полученные данные позволили сформировать как энтальпийную, так и свободноэнергетическую шкалу. Мультипликативная модель по отношению к ΔG была представлена в виде:

$$\Delta G_{ij} = |\Delta G_{11}| C_i C_j \quad (26)$$

где ΔG_{11} – свободная энергия Гиббса образования Н-комплекса состава 1:1 в CCl_4 при 298 К между стандартным протонодонором (фенол) и стандартным протоноакцептором (диэтиловый эфир).

Проведя систематическую работу по сбору, анализу и обработке термодинамических характеристик реакций образования для нескольких тысяч Н-комплексов, нам удалось сформировать базу данных и получить возможность создать статистически достоверные, полные термодинамические шкалы Н-связи (а именно, шкалы как для ΔH , так и для ΔG), основанные на мультипликативном подходе и ориентированные на использование в QSAR моделировании. Для этого из полученных данных были выбраны реакции, удовлетворяющие следующим условиям:

- 1) Н-комплекс состава 1:1, с одним водородным мостиком;
- 2) растворитель CCl_4 ;
- 3) для каждой реакции имеются значения как ΔH , так и ΔG ;
- 4) ΔH и ΔG рассчитаны на основе прямых экспериментальных данных и соответствуют температуре 298 К.

Общее число отобранных реакций составило 936 и все эти реакции были включены в дальнейшую процедуру обработки. Расчет энтальпийных и свободноэнергетических дескрипторов Н-связи был выполнен с использованием следующих уравнений:

$$\Delta H = k_1 E_a E_d + k_0 \quad (27)$$

$$\Delta G = k_1 C_a C_d + k_0 \quad (28)$$

где подстрочная буква «d» означает Н-донор, а буква «a» – Н-акцептор.

Для расчета E и C дескрипторов необходима фиксация величин для одного Н-донора и одного Н-акцептора. Эти соединения должны удовлетворять двум требованиям: 1) они должны часто встречаться в нашей базе данных в качестве одного из участников при образовании Н-комплекса и 2) они должны быть весьма «сильными», чтобы обеспечить получение надежных экспериментальных величин. Несомненно, фенол, в качестве стандартного Н-донора, удовлетворяет этим условиям: он встречается 119 раз и обеспечивает высокие значения ΔH и ΔG . Что касается стандартного Н-акцептора, то используемый ранее диэтиловый эфир плохо подходит для этой роли: он редко встречается в нашей БД и с ним трудно работать экспериментально. Поэтому в качестве протоноакцепторного стандарта мы выбрали гексаметилфосфортриамид (ГМФТА), встречающийся 88 раз и удобный в экспериментальном отношении. Для удобства мы решили сделать шкалы с интервалом изменения дескрипторов от 0.00 до ($\cong$ -5.00) для Н-доноров и от 0.00 до ($\cong$ 5.00) для Н-акцепторов. Принимая во внимание наши предыдущие шкалы, чтобы удовлетворить этим интервалам, величина энтальпийного дескриптора была выбрана равной -2.50 для фенола и 2.50 для ГМФТА, а свободноэнергетического дескриптора -2.50 и 4.00 для фенола и ГМФТА соответственно.

Для дальнейшей обработки была сформирована матрица данных, состоящая из 163 колонок (протонодоноры) и 195 строк (протоноакцепторы). Алгоритм расчета включал в себя следующие шаги:

- 1) найти протоноакцепторы, которые имеют реакции с фенолом;
- 2) найти протонодоноры, которые имеют реакции с этими Н-акцепторами;
- 3) сконструировать и решить систему уравнений для каждой пары реакций, образованной фенолом и данным Н-донором с тем же самым протоноакцептором.

Например, величины E и C для этанола могут быть рассчитаны из следующей системы уравнений:

$$\Delta H_{\text{фп}} = k_1 E_{\text{д}}^{\text{фенол}} E_{\text{а}}^{\text{пиридин}} + k_0 \quad (29)$$

$$\Delta H_{\text{эп}} = k_1 E_{\text{д}}^{\text{этанол}} E_{\text{а}}^{\text{пиридин}} + k_0 \quad (30)$$

$$\Delta G_{\text{фп}} = k_1 C_{\text{д}}^{\text{фенол}} C_{\text{а}}^{\text{пиридин}} + k_0 \quad (31)$$

$$\Delta G_{\text{эп}} = k_1 C_{\text{д}}^{\text{этанол}} C_{\text{а}}^{\text{пиридин}} + k_0 \quad (32)$$

где $\Delta H_{\text{фп}}$ и $\Delta G_{\text{фп}}$ – энтальпия и свободная энергия Гиббса образования комплекса между фенолом и пиридином, $\Delta H_{\text{эп}}$ и $\Delta G_{\text{эп}}$ – то же самое для этанола и пиридина. Для большинства Н-доноров можно найти больше, чем одну систему в базе данных, и решать их независимо. В этом случае результат представляет собой среднюю величину из всех полученных решений. В случае, если было невозможно комбинировать такие наборы уравнений для некоторых соединений (они не имели реакций с подходящими Н-акцепторами, которые реагировали с фенолом), мы повторяли представленный выше алгоритм используя дескрипторы, полученные при первом расчете, таким же образом, как были использованы дескрипторы фенола. Протоноакцепторные дескрипторы рассчитывались с использованием того же самого алгоритма, используя ГМФТА в качестве стандартного Н-акцептора.

Таблица 5

Величины энтальпийных Н-донорных (E_d), Н-акцепторных (E_a) и свободноэнергетических Н-донорных (C_d), Н-акцепторных (C_a) дескрипторов органических соединений (s – стандартное отклонение, n – число соединений, на основе которых были рассчитаны дескрипторы)

№	Соединение	E_d	s	C_d	s	n	E_a	s	C_a	s	n
1	<chem>CHCl3</chem>	-0.83	0.01	-0.80	0.05	2	-	-	-	-	-
2	<chem>CHBr3</chem>	-0.84	-	-0.80	-	1	-	-	-	-	-
3	<chem>CBr3CHBr2</chem>	-0.67	-	-0.64	-	1	-	-	-	-	-
4	<chem>CH2Br2</chem>	-0.51	-	-0.51	-	1	-	-	-	-	-
5	<chem>C6H5CHBr2</chem>	-0.51	-	-0.51	-	1	-	-	-	-	-
6	<chem>4-NO2-C6H4CHBr2</chem>	-1.17	-	-1.07	-	1	-	-	-	-	-
7	<chem>CHF2CN</chem>	-1.34	-	-1.25	-	1	-	-	-	-	-
8	<chem>CHCl2CN</chem>	-1.68	-	-1.43	-	1	-	-	-	-	-
9	<chem>CHBr2CN</chem>	-1.53	0.13	-1.40	0.05	3	-	-	-	-	-
10	<chem>CH2ClCN</chem>	-1.17	-	-1.02	-	1	-	-	-	-	-
11	<chem>C6H5CH(CN)CO2C2H5</chem>	-0.67	-	-0.68	-	1	-	-	-	-	-
12	<chem>CH2(CN)2</chem>	-1.89	0.26	-1.87	0.10	3	-	-	-	-	-
13	<chem>CHBr(CN)2</chem>	-2.52	-	-2.05	-	1	-	-	-	-	-
14	<chem>CHBr2NO2</chem>	-1.34	-	-1.24	-	1	-	-	-	-	-
15	<chem>C6H5CH(NO2)2</chem>	-1.57	0.39	-1.52	0.01	2	-	-	-	-	-
16	<chem>C6H5CH(CN)2</chem>	-1.38	0.13	-1.59	0.10	2	-	-	-	-	-
17	<chem>CHCl2CON(CH3)2</chem>	-0.67	-	-0.71	-	1	-	-	-	-	-
18	<chem>CHCl2CO2CH3</chem>	-0.51	-	-0.51	-	1	-	-	-	-	-
19	<chem>Cy-C3HCl5</chem>	-0.51	-	-0.47	-	1	-	-	-	-	-
20	<chem>CBr2=CHBr</chem>	-0.34	-	-0.23	-	1	-	-	-	-	-
21	<chem>CCl2=CHCl</chem>	-0.34	-	-0.27	-	1	-	-	-	-	-
22	<chem>(E)-N#CCH=CHCN</chem>	-1.68	-	-1.34	-	1	-	-	-	-	-
23	<chem>C2H5OCH=C(CN)2</chem>	-1.68	-	-1.39	-	1	-	-	-	-	-
24	<chem>(CH3)2C=CHCN</chem>	-0.34	-	-0.26	-	1	-	-	-	-	-
25	<chem>(E)-C6H5CH=CHCN(\alpha)</chem>	-0.67	-	-0.66	-	1	-	-	-	-	-

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
26	(E)-C ₆ H ₅ CH=CHCN(β)	-0.67	-	-0.60	-	1	-	-	-	-	-
27	(E)-C ₆ H ₅ CH=CHNO ₂ (α)	-0.67	-	-0.63	-	1	-	-	-	-	-
28	(E)-C ₆ H ₅ CH=CHNO ₂ (β)	-0.67	-	-0.67	-	1	-	-	-	-	-
29	1,2,4,5-Cl ₄ -C ₆ H ₂	-0.34	-	-0.52	-	1	-	-	-	-	-
30	C ₆ HCl ₅	-0.34	-	-0.00	-	1	-	-	-	-	-
31	3-NO ₂ -1,2,4,5-Cl ₄ -C ₆ H	-0.51	-	-0.43	-	1	-	-	-	-	-
32	1,2,4,5-F ₄ -C ₆ H ₂	-0.51	-	-0.40	-	1	-	-	-	-	-
33	C ₅ H ₅ C≡CH	-0.34	-	-0.12	-	1	-	-	-	-	-
34	CH ₂ BrC≡CH	-0.51	-	-0.45	-	1	-	-	-	-	-
35	C ₂ H ₅ O ₂ C≡CH	-0.67	-	-0.74	-	1	-	-	-	-	-
36	NCC≡CH	-1.34	-	-1.28	-	1	-	-	-	-	-
37	C ₆ H ₆	-	-	-	-	-	0.50	-	1.35	-	1
38	CH ₃ -C ₆ H ₅	-	-	-	-	-	0.52	-	1.60	-	1
39	C ₂ H ₅ -C ₆ H ₅	-	-	-	-	-	0.54	-	1.69	-	1
40	C ₃ H ₇ -C ₆ H ₅	-	-	-	-	-	0.54	-	1.69	-	1
41	C ₄ H ₉ -C ₆ H ₅	-	-	-	-	-	0.54	-	1.77	-	1
42	1,2-(CH ₃) ₂ -C ₆ H ₄	-	-	-	-	-	0.53	-	1.77	-	1
43	1,3-(CH ₃) ₂ -C ₆ H ₄	-	-	-	-	-	0.54	-	1.73	-	1
44	1,4-(CH ₃) ₂ -C ₆ H ₄	-	-	-	-	-	0.55	-	1.81	-	1
45	1,3,5-(CH ₃) ₃ -C ₆ H ₃	-	-	-	-	-	0.60	-	2.07	-	1
46	1,2,4,5-(CH ₃) ₄ -C ₆ H ₂	-	-	-	-	-	0.61	-	2.28	-	1
47	(CH ₃) ₅ -C ₆ H	-	-	-	-	-	0.64	-	2.51	-	1
48	(CH ₃) ₆ -C ₆	-	-	-	-	-	0.66	-	2.70	-	1
49	(CF ₃) ₃ C-NH ₂	-1.59	0.28	-1.44	0.02	2	-	-	-	-	-
50	CH ₃ -C(S)-NH ₂	-1.35	-	-2.47	-	1	-	-	-	-	-
51	C ₆ H ₅ -NH ₂	-0.84	-	-0.46	-	1	-	-	-	-	-
52	4-NO ₂ -C ₆ H ₄ -NH ₂	-1.13	-	-1.80	-	1	-	-	-	-	-
53	C(O)H-NH-C(CH ₃) ₃	-2.01	-	-0.64	-	1	-	-	-	-	-
54	CH ₃ -NH-NO ₂	-2.24	0.29	-2.34	0.37	5	-	-	-	-	-
55	C ₃ H ₇ -NH-NO ₂	-2.07	0.25	-2.25	0.32	5	-	-	-	-	-
56	C ₂ H ₅ -O-C(O)-NH-NO ₂	-2.61	0.12	-2.66	0.08	4	-	-	-	-	-
57	NC-(CH ₂) ₂ -NH-NO ₂	-2.46	0.25	-2.94	0.36	5	-	-	-	-	-
58	C ₄ H ₉ -NH-NO ₂	-2.14	0.19	-2.26	0.30	5	-	-	-	-	-
59	(CH ₃) ₃ C-NH-NO ₂	-2.22	-	-2.24	-	1	-	-	-	-	-

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
60	(NO ₂) ₃ C-(CH ₂) ₂ -NH-NO ₂	-2.73	0.32	-3.13	0.40	5	-	-	-	-	-
61	C ₆ H ₁₁ -NH-NO ₂	-2.09	0.16	-2.24	0.10	4	-	-	-	-	-
62	CH ₃ -C(O)-NH-C ₆ H ₅	-2.09	-	-1.95	-	1	-	-	-	-	-
63	CH ₃ -C(O)-NH-(4-Br-C ₆ H ₄)	-2.63	-	-2.12	-	1	-	-	-	-	-
64	CH ₃ -C(O)-NH-(4-OCH ₃ -C ₆ H ₄)	-1.60	-	-1.88	-	1	-	-	-	-	-
65	4-OCH ₃ -C ₆ H ₄ -SO ₂ -NH-NO ₂	-2.95	-	-3.06	-	1	-	-	-	-	-
66	4-OCH ₃ -C ₆ H ₄ -NH-C ₆ H ₅	-0.89	-	-0.91	-	1	-	-	-	-	-
67	(4-OCH ₃ -C ₆ H ₄) ₂ NH	-1.24	-	-2.73	-	1	-	-	-	-	-
68	(C ₆ H ₅) ₂ NH	-1.36	-	-1.07	-	1	0.72	0.04	1.36	0.06	2
69	(4-CH ₃ -C ₆ H ₄) ₂ NH	-0.77	-	-0.85	-	1	-	-	-	-	-
70	(4-Br-C ₆ H ₄) ₂ NH	-1.49	-	-1.54	-	1	0.72	0.04	1.62	0.07	2
71	N-фенил-1-нафтиламин	-0.89	-	-0.39	-	1	1.13	0.06	1.21	0.06	2
72	N-фенил-2-нафтиламин	-1.71	-	-1.14	-	1	0.93	0.05	1.41	0.06	2
73	Пиррол	-1.34	0.23	-1.67	0.22	23	1.29	-	1.32	-	1
74	Индол	-1.55	0.20	-1.90	0.18	19	-	-	-	-	-
75	Карбазол	-1.72	0.20	-2.04	0.19	16	-	-	-	-	-
76	1,4-дiazобикакло[2,2,2]октан	-	-	-	-	-	2.42	0.55	4.99	2.78	2
77	Пиперидин	-	-	-	-	-	2.69	-	3.01	-	1
78	1-метилпиперидин	-	-	-	-	-	2.09	0.15	2.59	0.27	3
79	Пиперазин	-	-	-	-	-	2.86	-	3.10	-	1
80	N,N'-диметилпиперазин	-	-	-	-	-	4.01	4.12	4.21	2.35	3
81	3,6-диметилпиперазин	-	-	-	-	-	2.41	-	2.93	-	1
82	(C ₂ H ₅) ₃ N	-	-	-	-	-	3.50	0.72	2.21	0.55	31
83	(C ₄ H ₉) ₃ N	-	-	-	-	-	4.05	0.48	3.65	0.36	2
84	(C ₄ H ₉) ₂ NH	-	-	-	-	-	5.21	0.68	3.65	0.36	3
85	(C ₃ H ₇) ₃ N	-	-	-	-	-	2.29	0.65	2.81	0.89	5
86	Пирролидин	-	-	-	-	-	2.79	-	3.14	-	1
87	Пиридин	-	-	-	-	-	2.37	0.34	2.61	0.28	12
88	2-метилпиридин	-	-	-	-	-	2.14	-	2.65	-	1
89	4-метилпиридин	-	-	-	-	-	2.20	-	2.68	-	1
90	3-метилпиридин	-	-	-	-	-	2.60	0.53	2.32	0.32	23
91	2,4-диметилпиридин	-	-	-	-	-	2.36	-	2.74	-	1
92	3,5-диметилпиридин	-	-	-	-	-	2.77	0.57	2.44	0.34	23
93	2,6-диметилпиридин	-	-	-	-	-	2.24	0.32	2.13	0.10	2
94	2,4,6-триметилпиридин	-	-	-	-	-	3.96	2.10	3.11	0.29	2

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
95	Пиридазин	-	-	-	-	-	2.28	-	2.35	-	1
96	Пиримидин	-	-	-	-	-	2.14	-	2.08	-	1
97	Пиразин	-	-	-	-	-	2.02	-	1.98	-	1
98	N-этилпиразол	-	-	-	-	-	1.61	-	2.57	-	1
99	N-этил-5-метилпиразол	-	-	-	-	-	1.54	-	2.75	-	1
100	N-этил-3-метилпиразол	-	-	-	-	-	1.48	-	2.67	-	1
101	N-этил-3,5-диметилпиразол	-	-	-	-	-	1.51	-	2.97	-	1
102	N-винилпиразол	-	-	-	-	-	1.17	-	1.95	-	1
103	N-винил-3-метилпиразол	-	-	-	-	-	1.31	-	2.17	-	1
104	N-винил-5-метилпиразол	-	-	-	-	-	1.38	-	2.19	-	1
105	N-винил-3,5-диметилпиразол	-	-	-	-	-	1.58	-	2.15	-	1
106	N,N,2,2-тетраметил-2Н-азири-3амин	-	-	-	-	-	2.41	0.10	3.06	0.06	4
107	N-метилимидазол	-	-	-	-	-	2.28	-	3.12	-	1
108	N,2-диэтилимидазол	-	-	-	-	-	2.35	-	3.37	-	1
109	N-метилбензимидазол	-	-	-	-	-	2.18	-	2.89	-	1
110	N-фенилимидазол	-	-	-	-	-	2.01	-	2.89	-	1
111	1-этилиндазол	-	-	-	-	-	1.31	-	2.62	-	1
112	1-винилиндазол	-	-	-	-	-	0.91	-	2.27	-	1
113	Пиридин-d ₅	-	-	-	-	-	3.75	-	2.99	-	1
114	C ₆ H ₅ -CH=N-CH ₃	-	-	-	-	-	1.91	0.07	1.95	0.04	2
115	CH ₃ -CN	-	-	-	-	-	1.69	0.27	1.55	0.15	6
116	Цианоциклопентан	-	-	-	-	-	1.92	-	1.60	-	1
117	CCl ₃ -CN	-	-	-	-	-	0.54	-	0.43	-	1
118	C ₆ H ₅ -CN	-	-	-	-	-	1.73	-	1.47	-	1
119	4-F-C ₆ H ₄ -CN	-	-	-	-	-	1.56	-	1.34	-	1
120	3-CF ₃ -C ₆ H ₄ -CN	-	-	-	-	-	1.52	-	1.30	-	1
121	C ₆ F ₅ -CN	-	-	-	-	-	0.92	-	0.86	-	1
122	CH ₃ -S-CN	-	-	-	-	-	1.67	-	1.32	-	1
123	(C ₆ H ₅) ₂ P-CN	-	-	-	-	-	1.67	-	1.53	-	1
124	(CH ₃) ₃ Si-CN	-	-	-	-	-	1.87	-	1.65	-	1
125	C ₂ H ₅ -CN	-	-	-	-	-	1.11	-	1.48	-	1
126	(CH ₃) ₂ CH-CH ₂ -CN	-	-	-	-	-	2.01	-	1.54	-	1
127	C ₆ H ₅ -CH ₂ -CN	-	-	-	-	-	1.87	-	1.62	-	1
128	CH ₂ =CH-CN	-	-	-	-	-	0.74	-	1.26	-	1
129	CH ₂ Cl-CH ₂ -CN	-	-	-	-	-	1.82	-	1.37	-	1

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
130	CD ₃ -CN	-	-	-	-	-	1.21	0.23	1.68	0.20	7
131	CH ₃ -OH	-1.82	0.23	-1.35	0.33	7	1.73	-	1.93	-	1
132	CH ₃ -CH ₂ -OH	-1.59	-	-1.46	-	1	1.78	-	1.86	-	1
133	CF ₃ -CH ₂ -OH	-2.36	0.19	-2.49	0.10	15	-	-	-	-	-
134	CCl ₃ -CH ₂ -OH	-2.12	0.22	-2.14	0.06	17	-	-	-	-	-
135	CBr ₃ -CH ₂ -OH	-2.12	0.28	-2.05	0.06	15	-	-	-	-	-
136	CF(Cl)(NO ₂)-CH ₂ -OH	-2.18	-	-3.15	-	1	-	-	-	-	-
137	CH(CH ₃) ₂ -OH	-1.53	-	-1.37	-	1	1.81	-	1.76	-	1
138	CF ₃ -CF ₂ -CH ₂ -OH	-2.82	-	-2.38	-	1	-	-	-	-	-
139	CF ₃ -CH(CF ₃)-OH	-2.78	-	-3.13	-	1	-	-	-	-	-
140	CCl ₃ -CH(CCl ₃)-OH	-2.69	-	-2.70	-	1	-	-	-	-	-
141	CF(Cl)(NO ₂)-CH(CH ₃)-OH	-2.27	-	-2.53	-	1	-	-	-	-	-
142	n-C ₄ H ₉ -OH	-1.49	0.09	-1.26	0.13	3	1.80	-	1.82	-	1
143	(CH ₃) ₂ CH-CH ₂ -OH	-1.56	-	-1.08	-	1	-	-	-	-	-
144	C(CH ₃) ₃ -OH	-1.28	0.18	-0.97	0.46	2	1.87	-	1.89	-	1
145	CF ₃ -(CF ₂) ₂ -CH ₂ -OH	-2.62	0.47	-2.46	0.30	2	-	-	-	-	-
146	CF ₃ -C(CH ₃) ₂ -OH	-1.99	-	-1.93	-	1	-	-	-	-	-
147	CH ₃ -C(CF ₃) ₂ -OH	-2.50	-	-2.74	-	1	-	-	-	-	-
148	C(CF ₃) ₃ -OH	-3.25	-	-3.48	-	1	-	-	-	-	-
149	CF ₃ -C(CCl ₃)(CF ₃)-OH	-2.73	-	-3.13	-	1	-	-	-	-	-
150	CCl ₃ -C(CH ₃) ₂ -OH	-1.71	-	-1.65	-	1	-	-	-	-	-
151	CF(Cl)(NO ₂)-CH(C ₂ H ₅)-OH	-2.25	-	-2.80	-	1	-	-	-	-	-
152	(CH ₃) ₂ CH-(CH ₂) ₂ -OH	-	-	-	-	-	1.81	-	1.72	-	1
153	C ₂ H ₅ -C(CH ₃)(C ₂ H ₅)-OH	-1.38	-	-1.17	-	1	-	-	-	-	-
154	CF(Cl)(NO ₂)-CH(C ₃ H ₇)-OH	-2.25	-	-2.52	-	1	-	-	-	-	-
155	CF ₃ -(CF ₂) ₂ -CHF-CH ₂ -OH	-2.57	0.46	-1.99	0.24	2	-	-	-	-	-
156	CF ₃ -(CF ₂) ₃ -CHF-(CH ₂) ₂ -OH	-2.44	0.44	-1.64	0.20	2	-	-	-	-	-
157	CF ₃ -(CF ₂) ₆ -CH ₂ -OH	-2.33	0.42	-1.40	0.17	2	-	-	-	-	-
158	CF(Cl)(NO ₂)-CH(C ₆ H ₅)-OH	-2.01	-	-2.79	-	1	-	-	-	-	-
159	CF(Cl)(NO ₂)-CH(4-F-C ₆ H ₄)-OH	-2.05	-	-3.09	-	1	-	-	-	-	-
160	CF(Cl)(NO ₂)-CH(2-F-C ₆ H ₄)-OH	-2.14	-	-2.81	-	1	-	-	-	-	-
161	CF(Cl)(NO ₂)-CH(2-OCH ₃ -C ₆ H ₄)-OH	-2.37	-	-2.46	-	1	-	-	-	-	-
162	SH-(CH ₂) ₂ -OH	-2.22	0.40	-1.60	0.20	2	-	-	-	-	-
163	C ₆ H ₅ -OH	-2.50	-	-2.50	-	-	-	-	-	-	-
164	2-CH ₃ -C ₆ H ₄ -OH	-1.60	-	-2.33	-	1	-	-	-	-	-

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
165	2-Cl-C ₆ H ₄ -OH	-2.20	0.39	-1.67	0.18	2	-	-	-	-	-
166	2-OCH ₃ -C ₆ H ₄ -OH	-1.36	0.18	-1.21	0.21	3	-	-	-	-	-
167	3-CH ₃ -C ₆ H ₄ -OH	-2.06	-	-2.42	-	1	-	-	-	-	-
168	3-F-C ₆ H ₄ -OH	-2.73	-	-2.79	-	1	-	-	-	-	-
169	3-Cl-C ₆ H ₄ -OH	-2.71	0.12	-2.84	0.03	3	-	-	-	-	-
170	3-Br-C ₆ H ₄ -OH	-2.84	0.17	-2.91	0.04	4	-	-	-	-	-
171	3-OH-C ₆ H ₄ -OH	-3.05	0.66	-3.13	0.10	3	-	-	-	-	-
172	3-NO ₂ -C ₆ H ₄ -OH	-3.08	0.25	-3.39	0.17	6	-	-	-	-	-
173	4-CH ₃ -C ₆ H ₄ -OH	-2.36	0.22	-2.40	0.06	24	-	-	-	-	-
174	4-(s-C ₄ H ₉)-C ₆ H ₄ -OH	-2.41	0.19	-2.40	0.03	16	-	-	-	-	-
175	4-(t-C ₄ H ₉)-C ₆ H ₄ -OH	-2.58	0.26	-2.43	0.14	2	-	-	-	-	-
176	4-F-C ₆ H ₄ -OH	-2.42	0.29	-2.68	0.13	4	-	-	-	-	-
177	4-Cl-C ₆ H ₄ -OH	-2.65	0.23	-2.81	0.08	30	-	-	-	-	-
178	4-Br-C ₆ H ₄ -OH	-2.70	0.24	-2.83	0.08	28	-	-	-	-	-
179	4-I-C ₆ H ₄ -OH	-2.73	0.25	-2.87	0.05	19	-	-	-	-	-
180	4-OCH ₃ -C ₆ H ₄ -OH	-2.43	0.27	-2.41	0.08	26	-	-	-	-	-
181	4-CH ₃ C(O)-C ₆ H ₄ -OH	-2.80	0.23	-3.05	0.09	17	-	-	-	-	-
182	4-CN-C ₆ H ₄ -OH	-3.10	0.23	-3.37	0.10	7	-	-	-	-	-
183	4-NO ₂ -C ₆ H ₄ -OH	-3.17	0.31	-3.50	0.12	20	-	-	-	-	-
184	2,4-Cl ₂ -C ₆ H ₃ -OH	-2.13	-	-2.14	-	1	-	-	-	-	-
185	2,4-(NO ₂) ₂ -C ₆ H ₃ -OH	-1.99	1.80	-1.12	0.34	4	-	-	-	-	-
186	2,5-Cl ₂ -C ₆ H ₃ -OH	-2.15	-	-2.36	-	1	-	-	-	-	-
187	2,6-(CH ₃) ₂ -C ₆ H ₃ -OH	-1.60	0.22	-1.72	0.12	8	-	-	-	-	-
188	2-CH ₃ -6-(t-C ₄ H ₉)-C ₆ H ₃ -OH	-1.66	0.06	-1.55	0.05	2	-	-	-	-	-
189	2,6-(i-C ₃ H ₇) ₂ -C ₆ H ₃ -OH	-1.63	-	-1.73	-	1	-	-	-	-	-
190	2,6-(t-C ₄ H ₉) ₂ -C ₆ H ₃ -OH	-1.04	-	-0.94	-	1	-	-	-	-	-
191	2,6-Cl ₂ -C ₆ H ₃ -OH	-2.24	-	-2.02	-	1	-	-	-	-	-
192	3,4-(CH ₃) ₂ -C ₆ H ₃ -OH	-2.19	0.20	-2.27	0.08	6	-	-	-	-	-
193	3,4-Cl ₂ -C ₆ H ₃ -OH	-2.96	0.26	-3.16	0.09	10	-	-	-	-	-
194	3-CF ₃ -4-NO ₂ -C ₆ H ₃ -OH	-3.70	0.85	-4.14	0.27	2	-	-	-	-	-
195	3,5-Cl ₂ -C ₆ H ₃ -OH	-3.20	0.33	-3.33	0.10	11	-	-	-	-	-
196	2,3,6-Cl ₃ -C ₆ H ₂ -OH	-2.31	-	-2.32	-	1	-	-	-	-	-
197	2,4,5-Cl ₃ -C ₆ H ₂ -OH	-2.58	0.34	-2.26	0.26	3	-	-	-	-	-
198	2,4,6-Cl ₃ -C ₆ H ₂ -OH	-2.25	0.35	-2.09	0.24	7	-	-	-	-	-
199	2,4,6-(t-C ₄ H ₉) ₃ -C ₆ H ₂ -OH	-1.76	0.87	-0.52	0.31	2	1.35	0.07	1.59	0.07	2

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
200	2,6-(C ₆ H ₅) ₂ -4-NO ₂ -C ₆ H ₂ -OH	-2.33	0.15	-2.02	0.02	3	-	-	-	-	-
201	2,6-(4-CH ₃ -C ₆ H ₄) ₂ -4-NO ₂ -C ₆ H ₂ -OH	-1.87	0.26	-1.77	0.03	3	-	-	-	-	-
202	2,6-(2-CH ₃ -C ₆ H ₄) ₂ -4-NO ₂ -C ₆ H ₂ -OH	-2.19	0.08	-2.06	0.03	3	-	-	-	-	-
203	2,6-(4-t-C ₄ H ₉ C ₆ H ₄) ₂ -4-NO ₂ -C ₆ H ₂ OH	-1.78	0.28	-1.71	0.05	3	-	-	-	-	-
204	2,6(2,6(CH ₃) ₂ C ₆ H ₄) ₂ -4-NO ₂ C ₆ H ₂ OH	-1.91	0.24	-1.97	0.04	3	-	-	-	-	-
205	3,4,5-Cl ₃ -C ₆ H ₂ -OH	-3.39	0.35	-3.52	0.15	8	-	-	-	-	-
206	2,3,4,6-Cl ₄ -C ₆ H-OH	-2.69	0.37	-2.37	0.21	2	-	-	-	-	-
207	C ₆ F ₅ -OH	-3.24	0.34	-3.47	0.34	4	-	-	-	-	-
208	C ₆ Cl ₅ -OH	-2.97	0.99	-2.53	0.18	5	-	-	-	-	-
209	C ₆ Br ₅ -OH	-2.08	0.14	-2.10	0.04	2	-	-	-	-	-
210	1-нафтол	-2.65	0.11	-2.59	0.03	18	-	-	-	-	-
211	2-нафтол	-2.61	0.16	-2.63	0.04	11	-	-	-	-	-
212	1-перфторнафтол	-3.14	0.28	-3.01	0.20	6	-	-	-	-	-
213	2-перфторнафтол	-3.07	0.33	-3.43	0.31	6	-	-	-	-	-
214	(CH ₃) ₃ C-O-OH	-2.05	-	-1.85	-	1	-	-	-	-	-
215	C ₆ H ₅ -C(CH ₃) ₂ -O-OH	-1.90	-	-1.67	-	1	-	-	-	-	-
216	4-Cl-C ₆ H ₄ -C(O)-O-OH	-1.98	0.46	-1.55	0.17	6	-	-	-	-	-
217	3-Cl-C ₆ H ₄ -C(O)-O-OH	-2.70	-	-1.61	-	1	-	-	-	-	-
218	4-t-C ₄ H ₉ -C ₆ H ₄ -O-OH	-2.40	-	-1.30	-	1	-	-	-	-	-
219	CH ₂ Cl-COOH	-3.43	0.12	-3.54	0.27	4	-	-	-	-	-
220	CHCl ₂ -COOH	-3.35	0.65	-4.29	0.52	8	-	-	-	-	-
221	CCl ₃ -COOH	-3.93	0.45	-4.78	0.60	10	-	-	-	-	-
222	C ₆ H ₅ -COOH	-2.81	0.56	-2.99	0.02	2	-	-	-	-	-
223	C ₆ Cl ₅ -COOH	-3.19	0.60	-3.38	0.11	2	-	-	-	-	-
224	C ₆ F ₅ -COOH	-3.66	0.47	-3.97	0.26	2	-	-	-	-	-
225	C ₂ H ₅ -O-C ₂ H ₅	-	-	-	-	-	1.78	0.11	1.72	0.03	2
226	C ₃ H ₇ -O-C ₃ H ₇	-	-	-	-	-	1.89	-	1.61	-	1
227	(CH ₃) ₂ CH-O-CH(CH ₃) ₂	-	-	-	-	-	1.99	-	1.80	-	1
228	C ₄ H ₉ -O-C ₄ H ₉	-	-	-	-	-	1.93	-	1.60	-	1
229	C ₆ H ₅ -O-CH ₃	-	-	-	-	-	1.07	0.18	1.08	0.10	6
230	4-CH ₃ -C ₆ H ₄ -O-CH ₃	-	-	-	-	-	1.01	-	1.14	-	1
231	4-F-C ₆ H ₄ -O-CH ₃	-	-	-	-	-	0.97	-	0.78	-	1
232	4-Cl-C ₆ H ₄ -O-CH ₃	-	-	-	-	-	0.77	-	0.96	-	1
233	Метилоксиран	-	-	-	-	-	1.34	-	1.70	-	1
234	Этилоксиран	-	-	-	-	-	1.53	-	1.66	-	1

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
235	2,2,3-триметилоксиран	-	-	-	-	-	1.45	-	1.91	-	1
236	Гексилоксиран	-	-	-	-	-	1.37	-	1.68	-	1
237	Хлорметилоксиран	-	-	-	-	-	0.64	-	1.50	-	1
238	Этенилоксиран	-	-	-	-	-	1.13	-	1.63	-	1
239	Тетрагидропиран	-	-	-	-	-	1.51	0.12	1.84	0.06	3
240	Тетрагидрофуран	-	-	-	-	-	1.94	0.31	2.09	0.26	10
241	Фуран	-	-	-	-	-	1.61	-	2.14	-	1
242	CH ₃ -O-(CH ₃) ₂ -O-CH ₃	-	-	-	-	-	1.85	-	2.02	-	1
243	4-ОСН ₃ -С ₆ Н ₄ -О-СН ₃	-	-	-	-	-	1.28	-	1.43	-	1
244	1,3-диоксолан	-	-	-	-	-	1.37	-	1.47	-	1
245	1,3-диоксан	-	-	-	-	-	1.84	0.52	1.52	0.08	2
246	1,4-диоксан	-	-	-	-	-	1.62	0.22	1.81	0.09	8
247	5,5-(CH ₃) ₂ -1,3-диоксан	-	-	-	-	-	1.84	0.52	1.51	0.09	2
248	1,3-диоксепан	-	-	-	-	-	1.50	-	1.63	-	1
249	2,4,6-(CH ₃) ₃ -триоксан	-	-	-	-	-	1.59	0.39	1.25	0.30	2
250	CH ₃ -C(O)-CH ₃	-	-	-	-	-	1.35	0.13	2.03	0.28	9
251	CH ₃ -C(O)-C ₂ H ₅	-	-	-	-	-	1.89	0.07	1.93	0.09	3
252	C ₆ H ₅ -C(O)H	-	-	-	-	-	1.11	0.11	1.68	0.22	6
253	CH ₃ -C(O)-C ₆ H ₅	-	-	-	-	-	1.28	0.06	1.76	0.03	5
254	4-N(CH ₃) ₂ -C ₆ H ₄ -C(O)H	-	-	-	-	-	1.67	0.11	2.26	0.08	7
255	C ₆ H ₅ -C(O)-C ₆ H ₅	-	-	-	-	-	1.18	0.12	1.71	0.05	5
256	CH ₃ -C(O)-O-CH ₃	-	-	-	-	-	1.35	0.14	1.60	0.09	3
257	CH ₃ -C(O)-O-C ₂ H ₅	-	-	-	-	-	1.15	-	1.77	-	1
258	CH ₃ -C(O)-O-CH ₂ -C ₆ H ₅	-	-	-	-	-	1.36	-	2.68	-	1
259	(C ₆ H ₅) ₂ C=N-CH ₂ -C(O)-O-C ₂ H ₅	-	-	-	-	-	1.39	0.02	2.13	0.05	2
260	CH ₃ -C(O)-NH-CH ₃	-	-	-	-	-	1.34	-	2.94	-	1
261	CH ₃ -C(O)-N(CH ₃) ₂	-	-	-	-	-	2.24	0.20	2.99	0.12	10
262	C ₂ H ₅ -C(O)-N(CH ₃) ₂	-	-	-	-	-	2.15	-	2.85	-	1
263	CH ₂ Cl-C(O)-N(CH ₃) ₂	-	-	-	-	-	1.58	-	2.42	-	1
264	C ₆ H ₅ -C(O)-N(CH ₃) ₂	-	-	-	-	-	1.81	-	2.85	-	1
265	N-(CH ₃)-пирролидин-2,5-дион	-	-	-	-	-	1.40	0.04	2.04	0.09	9
266	CH ₃ -C(O)-N(C ₆ H ₅) ₂	-	-	-	-	-	1.73	-	2.54	-	1
267	C ₆ H ₅ -C(O)-N(C ₆ H ₅) ₂	-	-	-	-	-	1.32	-	2.35	-	1
268	(CH ₃) ₂ N-C(O)-N(CH ₃) ₂	-	-	-	-	-	2.14	0.23	2.95	0.05	11
269	CH ₃ -C(O)-N(C ₂ H ₅) ₂	-	-	-	-	-	2.66	0.48	3.10	0.42	7

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
270	N,N-диэтил-никотинамид	-	-	-	-	-	1.73	0.16	2.80	0.06	3
271	N-метил-пирролидон-2	-	-	-	-	-	1.94	0.13	3.09	0.03	9
272	Метилникотинат	-	-	-	-	-	1.53	0.02	2.17	0.07	3
273	2,6-диметилпирон	-	-	-	-	-	2.03	0.25	2.74	0.23	2
274	CH ₃ -C(O)-NH-C ₄ H ₉	-	-	-	-	-	2.39	0.06	2.79	0.12	2
275	CH ₃ -C(O)-NH-CH(CH ₃) ₂	-	-	-	-	-	2.20	0.23	2.66	1.43	4
276	3-Cl-C ₆ H ₄ -C(O)-O-OH	-	-	-	-	-	3.05	0.55	3.24	0.44	7
277	4-C ₄ H ₉ -C ₆ H ₄ -C(O)-O-OH	-	-	-	-	-	2.70	0.49	2.60	0.35	7
278	1-(4-CH ₃ -C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.79	0.31	2.60	0.28	58
279	1-(4-C ₂ H ₅ -C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.73	0.28	2.62	0.29	58
280	1-(3-CH ₃ -C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.73	0.59	2.56	0.28	58
281	1-(C ₆ H ₅)-2-пирролидинон	-	-	-	-	-	1.72	0.33	2.54	0.28	58
282	1-(3-OCH ₃ -C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.62	0.27	2.50	0.27	58
283	1-(4-OCH ₃ -C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.79	0.28	2.70	0.30	58
284	1-(4-Cl-C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.48	0.24	2.25	0.24	58
285	1-(3-Cl-C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.47	0.24	2.20	0.24	58
286	1-(2-CH ₃ -C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	2.14	0.45	2.95	0.33	58
287	1-(2-Cl-C ₆ H ₄)-2-пирролидинон	-	-	-	-	-	1.88	0.32	2.82	0.31	58
288	N,N'-диметилпиримидин-дион-2,4	-	-	-	-	-	1.87	0.07	2.43	0.08	5
289	2-оксотетрагидротиофен	-	-	-	-	-	1.26	0.09	1.71	0.02	9
290	Циклогексанон	-	-	-	-	-	1.44	0.10	2.06	0.09	7
291	(CD ₃) ₂ CO	-	-	-	-	-	2.09	-	1.63	-	1
292	2-Фуральдегид	-	-	-	-	-	1.42	0.33	1.63	0.04	3
293	(CH ₃) ₂ N-C(O)H	-	-	-	-	-	1.98	0.23	2.82	0.11	11
294	CH ₃ -C(O)-S-CH ₃	-	-	-	-	-	1.09	0.06	1.55	0.03	3
295	(C ₆ H ₅) ₃ PO	-	-	-	-	-	2.54	0.17	3.63	0.08	5
296	(CH ₃ O) ₃ PO	-	-	-	-	-	2.02	0.06	3.00	0.09	5
297	(C ₄ H ₉ O) ₃ PO	-	-	-	-	-	2.35	0.43	3.39	0.46	7
298	(C ₆ H ₅ O) ₃ PO	-	-	-	-	-	1.76	0.09	2.46	0.08	5
299	(C ₄ H ₉ O) ₂ P(O)-NH-C ₄ H ₉	-	-	-	-	-	2.21	-	3.57	-	1
300	(C ₄ H ₉ O)P(O)(NH-C ₄ H ₉) ₂	-	-	-	-	-	2.66	-	3.99	-	1
301	(NH-C ₄ H ₉) ₃ PO	-	-	-	-	-	2.95	-	4.62	-	1
302	((CH ₃) ₂ N) ₃ PO	-	-	-	-	-	2.50	-	4.00	-	-
303	(CH ₃ O) ₂ P(O)-CH ₃	-	-	-	-	-	2.02	0.08	3.30	0.11	2
304	(C ₂ H ₅ O) ₂ P(O)-CCl ₃	-	-	-	-	-	1.47	-	1.89	-	1

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	n
305	(C ₂ H ₅ O) ₃ PO	-	-	-	-	-	1.96	-	3.16	-	1
306	(C ₂ H ₅ O) ₂ P(O)-CH ₃	-	-	-	-	-	1.89	-	2.83	-	1
307	(4-CH ₃ -C ₆ H ₄) ₂ SO	-	-	-	-	-	1.90	0.02	2.78	0.04	3
308	(4-Cl-C ₆ H ₄) ₂ SO	-	-	-	-	-	1.72	0.05	2.36	0.05	3
309	CH ₃ -S(O)-CH=CH ₂	-	-	-	-	-	1.85	0.12	2.84	0.09	3
310	CH ₂ =CH-S(O)-CH=CH ₂	-	-	-	-	-	1.86	0.17	2.49	0.09	2
311	(CH ₃) ₂ SO	-	-	-	-	-	2.02	0.24	3.08	0.09	19
312	(CD ₃) ₂ SO	-	-	-	-	-	2.43	0.27	3.35	0.13	9
313	(C ₆ H ₅) ₂ SO	-	-	-	-	-	1.75	0.10	2.60	0.04	13
314	(CH ₃) ₂ N-S(O)-CH ₃	-	-	-	-	-	1.85	0.10	2.90	0.08	13
315	(CH ₃) ₂ N-S(O)-C ₆ H ₅	-	-	-	-	-	1.79	0.11	2.67	0.05	13
316	(CH ₃) ₂ N-S(O)-(4-CH ₃ -C ₆ H ₄)	-	-	-	-	-	1.84	0.03	2.76	0.07	5
317	(C ₆ H ₅ -CH ₂) ₂ SO	-	-	-	-	-	1.99	0.13	2.97	0.04	5
318	(CH ₃) ₂ SO ₂	-	-	-	-	-	1.42	-	2.00	-	1
319	CH ₃ -NH-SO ₂ -CH ₃	-	-	-	-	-	1.23	0.09	1.97	0.05	9
320	(CH ₃) ₂ N-SO ₂ -CH ₃	-	-	-	-	-	1.32	0.05	2.01	0.08	13
321	(C ₆ H ₅) ₂ N-SO ₂ -CH ₃	-	-	-	-	-	1.12	-	1.90	-	1
322	(CH ₃) ₂ N-SO ₂ -N(CH ₃) ₂	-	-	-	-	-	1.40	-	2.02	-	1
323	C ₆ H ₅ -SO ₂ -N(CH ₃) ₂	-	-	-	-	-	1.21	0.10	1.93	0.08	13
324	C ₆ H ₅ -SO ₂ -N(C ₆ H ₅) ₂	-	-	-	-	-	1.06	-	1.81	-	1
325	(C ₆ H ₅) ₂ SO ₂	-	-	-	-	-	1.22	0.10	1.93	0.08	8
326	(CD ₃ O) ₂ SO ₂	-	-	-	-	-	1.07	-	0.94	-	1
327	(CH ₃) ₂ N-SO ₂ -(4-CH ₃ -C ₆ H ₄)	-	-	-	-	-	1.20	0.10	2.06	0.11	5
328	(C ₆ H ₅) ₂ SeO	-	-	-	-	-	3.46	0.53	3.36	0.40	22
329	C ₂ H ₅ -SH	-2.18	0.39	-1.44	0.18	2	-	-	-	-	-
330	C ₃ H ₇ -SH	-2.07	0.37	-1.18	0.14	2	-	-	-	-	-
331	C ₄ H ₉ -SH	-2.04	0.37	-0.87	0.11	2	-	-	-	-	-
332	C ₅ H ₁₁ -SH	-2.00	0.36	-0.56	0.07	2	-	-	-	-	-
333	C ₇ H ₁₅ -SH	-1.82	0.33	-0.06	0.01	2	-	-	-	-	-
334	CH ₃ -C(O)-SH	-1.45	0.32	-1.59	0.25	4	-	-	-	-	-
335	CH ₂ Cl-C(O)-SH	-1.71	0.09	-1.91	0.02	2	-	-	-	-	-
336	C ₄ H ₉ -S-C ₄ H ₉	-	-	-	-	-	1.20	-	1.46	-	1
337	Тетрагидроотиофен	-	-	-	-	-	1.01	-	1.10	-	1
338	Тиофен	-	-	-	-	-	1.05	-	0.56	-	1
339	(CH ₃) ₂ N-C(S)-CH ₃	-	-	-	-	-	1.30	0.07	1.91	0.02	9

Продолжение табл. 5

№	Соединение	E _d	s	C _d	s	n	E _a	s	C _a	s	N
340	(CH ₃) ₂ N-C(S)-N(CH ₃) ₂	-	-	-	-	-	1.41	0.10	2.01	0.04	9
341	N-метилпирролидин-2-тион	-	-	-	-	-	1.23	0.07	2.03	0.02	8
342	(C ₆ H ₅) ₃ PS	-	-	-	-	-	1.17	0.16	1.69	0.03	5
343	((CH ₃) ₂ N) ₃ PS	-	-	-	-	-	1.29	0.08	2.01	0.02	9
344	(C ₂ H ₅ O) ₂ P(S)-O-(4-NO ₂ -C ₆ H ₄)	-	-	-	-	-	2.22	-	0.69	-	1
345	(C ₆ H ₅) ₃ Pse	-	-	-	-	-	1.15	0.13	1.65	0.04	5
346	2,6-диметил-4-тиопирон	-	-	-	-	-	1.63	0.03	2.52	0.12	2
347	2,6-диметил-4-селенопирон	-	-	-	-	-	1.37	0.09	2.21	0.60	2

Величины энтальпийных и свободноэнергетических дескрипторов для 347 соединений представлены в табл. 5. Наилучшие уравнения для всех рассчитанных дескрипторов были следующими (ΔH , ΔG в кДж/моль):

$$\Delta H = 4.96 E_a E_d \quad (33)$$

$$\Delta G = 2.43 C_a C_d + 5.70 \quad (34)$$

Величины дескрипторов Н-связи органических соединений, полученные на основе уравнений (33) и (34), являются базой для создания единой энергетической шкалы водородной связи (рис. 4). Полная количественная оценка способности органического соединения к образованию Н-связи может быть выполнена на основе четырех дескрипторов: E_a , E_d , C_a , C_d . Из представленных данных ясно, что молекулярное окружение оказывает существенное влияние на величины дескрипторов активных центров. Также можно констатировать, что разные по природе химические соединения могут иметь одинаковую способность к образованию водородной связи.

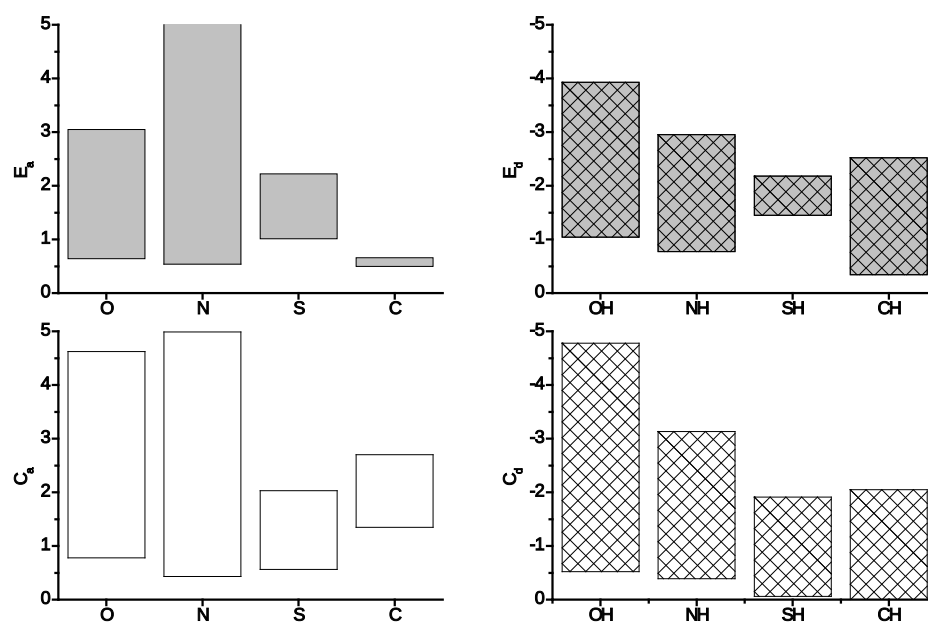


Рис. 4. Единая энергетическая шкала водородной связи

Необходимо подчеркнуть, что уравнения (33) и (34) отличаются друг от друга тем, что уравнение (34) содержит свободный член, а уравнение (33) – нет. Как уже было отмечено выше, мы предложили противоположные знаки для протонодонорных и протоноакцепторных дескрипторов: E_d и C_d имеют отрицательное значение, а E_a и C_a – положительное. Следовательно, положительные величины ΔH и ΔG могут быть получены только путем введения положительного свободного члена в уравнения (33) и (34). Таким образом, появление свободного члена является очевидным и остается только ответить на вопрос каким он должен быть по величине. В нашей базе данных, содержащей 936 значений как ΔH , так и ΔG в CCl_4 , было найдено 67 величин с положительной ΔG . Самая большая величина ΔG равна 5.87 кДж/моль для комплекса пентахлорбензол - ГМФТА в CCl_4 . Мы можем предположить, что возможны более высокие значения ΔG , например, в более полярных растворителях. Действительно, в работе [357] приведена величина ΔG для

образования Н-комплекса бутанол - метил-N-фенилкарбамат в бензоле равная 9.2 кДж/моль. Нужно отметить, что величины ΔG большие, чем 5.7 кДж/моль (что соответствует константе комплексообразования $K=0.1$), 1) близки к границе детектируемости и 2) характер Н-связи таких слабых комплексов проблематичен [358]. Поэтому мы ограничили нашу шкалу величиной 5.70 кДж/моль. В рамках рассматриваемого набора термодинамических данных мы имеем только один нулевой С-дескриптор (пентахлорбензол) и какие-либо более слабые Н-доноры будут иметь такую же величину С-дескриптора. Таким образом, мы принимаем, что свободная энергия Гиббса образования Н-комплекса с участием нулевого С-дескриптора имеет постоянное значение 5.70 кДж/моль. Аналогичные рассуждения могут быть применены и по отношению к ΔH . Однако, мы не нашли в нашей базе данных положительных величин энтальпии Н-связи.

Для того, чтобы критически оценить приведенный выше подход, мы использовали прямой тест. Так мы проверили поведение статистических критериев $\Delta H_{\text{эксп}} \cdot \Delta H_{\text{расч}}$ и $\Delta G_{\text{эксп}} \cdot \Delta G_{\text{расч}}$ для уравнений (33), (34) с величиной свободного члена в интервале от -20 кДж/моль до 20 кДж/моль. Тест показал следующее поведение критериев при изменении свободного члена в уравнении (34): уменьшение свободного члена ниже 5.70 кДж/моль приводило к уменьшению статистических характеристик (r^2 , F), но эти характеристики оставались приблизительно постоянными, если свободный член возрастал выше этой величины. Таким образом, это подтверждало наш выбор величины свободного члена для уравнения (34). Та же самая процедура тестирования, примененная к энтальпии, продемонстрировала достаточно существенное ухудшение статистических характеристик для отрицательных значений свободного члена, но приблизительно постоянные величины для положительных значений.

Используя полученные нами величины дескрипторов для 163 протонодоноров и 195 протоноакцепторов, можно оценить значения энтальпии

и свободной энергии Гиббса для $163 \times 195 = 31785$ реакций на основе уравнений (33) и (34). Для имеющихся в нашем распоряжении 936 экспериментальных и рассчитанных величин ΔH и ΔG были получены следующие соотношения:

$$\Delta H_{\text{расч}} = -0.27(\pm 0.45) + 1.00(\pm 0.02) \Delta H_{\text{эксп}} \quad (35)$$

$$n=936; r=0.954; s=2.70; F=9553$$

$$\Delta G_{\text{расч}} = -0.07(\pm 0.12) + 1.00(\pm 0.01) \Delta G_{\text{эксп}} \quad (36)$$

$$n=936; r=0.984; s=1.11; F=28556$$

Статистические характеристики уравнений (35) и (36) свидетельствуют о том, что дескрипторы в табл. 5 дают возможность реальной оценки энтальпии и свободной энергии Гиббса Н-связи. При этом анализ остатков показал, что 842 из 936 рассчитанных величин ΔG имеют менее чем 10% отклонение от экспериментальных величин. 94 случая больших отклонений были внимательно проанализированы и было установлено, что часть этих отклонений, как представляется, были обусловлены экспериментальными ошибками. Экспериментальные ошибки были выявлены путем сравнения сильно отклоняющихся величин с величинами в родственных реакциях. Например, нет причины предполагать, что CD_3CN и CH_3CN имеют значительную разницу в способности к образованию Н-связи, а дескрипторы, полученные для них, равны ($E_a=1.21$, $C_a=1.69$) и ($E_a=1.68$, $C_a=1.69$). Также для $\text{CH}_3\text{C}(\text{O})\text{CH}_3$ и $\text{CH}_3\text{C}(\text{O})\text{C}_2\text{H}_5$ имеются дескрипторы ($E_a=1.35$, $C_a=2.03$) и ($E_a=1.89$, $C_a=1.93$) соответственно. Другой замечательный пример состоит в том, что восемь величин ΔH и ΔG образования Н-комплекса между фенолом и пиридином в CCl_4 варьируют в интервале от -24.50 кДж/моль до -31.80 кДж/моль для ΔH и от -8.88 кДж/моль до -10.57 кДж/моль для ΔG .

Также интересно было проверить некоторые положения теории, предложенной в работе [359]. Главное из них состоит в возможности

конструирования параметрической шкалы свойств, определяемых основностью в зависимости от β величин соответствующих соединений. Мы обнаружили все пары реагентов, которые должны показывать «семейно зависимое» поведение. Действительно, было установлено, что 12 из 14 рассчитанных величин, давали положительное отклонение от экспериментальных величин (табл. 6) в соответствии с рассматриваемой теорией. Это означает, что некоторые классы химических соединений не очень хорошо укладываются в рамки простой параметрической модели. В работе [358] эта проблема решена путем конструирования параллельной шкалы кислотности Н-связи для «плохих» реакций.

Дополнительная информация о качестве шкал водородной связи может быть получена из анализа распределения Н-комплексов, из которого следует, что: 1) соединения из базы данных очень разнообразны по способности к образованию Н-связи, 2) большинство из них принимают участие в образовании Н-комплексов с величинами энтальпии и свободной энергии Гиббса в интервале от -8 кДж/моль до -44 кДж/моль для ΔH и от -10 кДж/моль до -30 кДж/моль для ΔG . При этом имеется достаточно данных, относящихся как к очень сильным, так и к очень слабым Н-комплексам.

Таблица 6

Экспериментальные и рассчитанные величины ΔG (кДж/моль) для некоторых реакций, которые должны показывать «семейно зависимое» поведение

Н-донор	Н-акцептор	$\Delta G_{\text{эксп}}$	$\Delta G_{\text{расч}}$	$\Delta G_{\text{эксп}} - \Delta G_{\text{расч}}$
Индол	Пиридин	-4.26	-6.30	2.04
Пиррол	Пиридин	-3.16	-4.86	1.70
Пиррол	2-CH ₃ -пиридин	-3.09	-5.06	1.97
Пиррол	3-CH ₃ -пиридин	-3.48	-3.69	0.21
Пиррол	4-CH ₃ -пиридин	-3.18	-5.17	1.99
Пиррол	2,4-(CH ₃) ₂ -пиридин	-3.60	-5.42	1.82
Пиррол	2,4,6-(CH ₃) ₃ -пиридин	-4.09	-6.92	2.83
t-Бутанол	2,6-(CH ₃) ₂ -пиридин	1.27	0.70	0.57
Пентахлорфенол	Тетрагидрофуран	-5.83	-7.17	1.34
Пентахлорфенол	Пиридин	-10.39	-10.33	-0.06
Пентахлорфенол	2,4,6-(CH ₃) ₃ -пиридин	-14.06	-13.45	-0.60
Пентахлорфенол	Пиридин-d ₅	-12.14	-12.72	0.58
Пентахлорфенол	2,4,6-(CH ₃) ₃ -триоксан	-0.47	-1.99	1.52
Пентахлорфенол	5,5-(CH ₃) ₂ -1,3-диоксан	-2.92	-3.60	0.68

Нами было проведено также сравнение свободноэнергетических дескрипторов с параметрами кислотности α_2^H [358] и основности β_2^H [360] Абрахама. Мы получили корреляции для Н-доноров:

$$C_d = -4.47(\pm 0.18) \alpha_2^H + 0.10(\pm 0.08) \quad (37)$$

$n=63$; $r=0.988$; $s=0.18$; $F=2469$

и для Н-акцепторов:

$$C_a = 3.98(\pm 0.67) \beta_2^H - 0.02(\pm 0.47) \quad (38)$$

$$n=11; r=0.976; s=0.17; F=178$$

Следует отметить, что дескрипторы представленной шкалы хорошо коррелируют с дескрипторами, полученными ранее [361]:

$$E_d = 2.29(\pm 0.16) E_i - 0.20(\pm 0.17) \quad (39)$$

$$n=20; r=0.990; s=0.11; F=899$$

Таким образом, сравнение представленных данных с предыдущими исследованиями по энтальпии и свободной энергии Гиббса ясно показали общность мультипликативного принципа в отношении термодинамики Н-связи органических соединений.

В результате проведенного исследования мы получили два ряда дескрипторов, а именно, E и C . Возникает вопрос: необходимы ли обе эти шкалы? Необходимость обоих дескрипторов была установлена с помощью C/E корреляций. Следующее соотношение было получено для Н-доноров:

$$C_d = 0.75(\pm 0.06) E_d - 0.52(\pm 0.13) \quad (40)$$

$$n=163; r=0.891; s=0.38; F=620$$

и для Н-акцепторов:

$$C_a = 0.61(\pm 0.10) E_a - 0.36(\pm 0.23) \quad (41)$$

$$n=195; r=0.664; s=0.53; F=152$$

НН-доноры:

$$C_d = 0.92(\pm 0.36) E_d - 0.24(\pm 0.67) \quad (42)$$

$n=27$; $r=0.726$; $s=0.55$; $F=28$

N-акцепторы:

$$C_a = 0.61(\pm 0.10) E_a - 0.36(\pm 0.23) \quad (43)$$

$n=195$; $r=0.664$; $s=0.53$; $F=152$

Для SH-доноров и S-акцепторов отсутствует коллинеарность между $E_d(E_a)$ и $C_d(C_a)$. В целом, данные в табл. 5 показывают тренд между величинами $E_d(E_a)$ и $C_d(C_a)$, который приблизительно одинаков для большинства соединений. Но существуют заметные отличия для отдельных соединений, которые могут быть обусловлены, например, стерическими затруднениями.

Таким образом для корректного описания H-связи при разработке новых биологически активных соединений, необходимо одновременное применение как энтальпийных, так и свободноэнергетических дескрипторов. Представляется, что использование E_d , E_a , и C_d , C_a возможно в двух областях: 1) оценка величин ΔH и ΔG при молекулярном моделировании субстрат-рецепторного взаимодействия и 2) QSAR моделирование.

Важным вопросом, требующим обсуждения, является вопрос, связанный с растворителем. Большинство исследований проводится с использованием тетрахлорметана, бензола и некоторых других неполярных и малополярных растворителей. Это легко объясняется с точки зрения экспериментальных удобств, таких как растворимость, спектральный шум и др. В нашей базе данных, кроме тетрахлорметана, фигурировало еще 29 других растворителей. Интересно было проверить поведение шкалы при изменении среды. Наиболее важный вопрос заключался в изучении постоянства величин дескрипторов в разных растворителях. Что касается энтальпийных протонодонорных и протоноакцепторных дескрипторов, то в работе Иогансена [353] было выдвинуто предположение, что при варьировании среды в уравнении (23)

должен меняться только коэффициент ΔH_{11} . Но эта гипотеза не может рассматриваться в качестве экспериментально подтвержденной, так как данные Иогансена были недостаточно надежными. Также имеется гипотеза Абрахама и сотр. относительно $\log K$ о том, что свободный член в уравнении (28) изменяется в зависимости от растворителя [362]. Собранные нами данные не позволили провести систематическое изучение поведения параметров в уравнениях (27), (28) при изменении среды. Мы имеем только несколько примеров. Один из них касается небольшого ряда соединений, для которых мы рассчитали дескрипторы в тетрахлорметане и хлороформе. Были получены следующие уравнения:

$$E_d^{\text{CCl}_4} = -0.59 + 0.73 E_d^{\text{CHCl}_3} \quad (44)$$

$$n=9; r=0.98; s=0.13; F=136$$

$$C_d^{\text{CCl}_4} = -0.21 + 0.90 C_d^{\text{CHCl}_3} \quad (45)$$

$$n=9; r=0.96; s=0.19; F=90$$

Уравнения (27), (28) для хлороформа были следующими (ΔH , ΔG в кДж/моль):

$$\Delta H = 3.32 E_a E_d \quad (46)$$

$$\Delta G = 1.93 C_a C_d + 5.70 \quad (47)$$

Мы также использовали данные из БД (C_a) и константы Н-связи в 1,1,1-трихлорэтаноле (K_{β}) из работы [363] для исследования их взаимосвязи:

$$C_a = 0.81(\pm 0.15) + 0.70(\pm 0.07) \log K_{\beta} \quad (48)$$

$$n=22; r=0.98; s=0.14; F=411$$

Представляется, что отношение дескрипторов Н-связи одного соединения относительно другого остается приблизительно постоянным в различных средах. Таким образом, энтальпийные и свободноэнергетические дескрипторы могут рассматриваться в качестве количественной меры способности органических соединений к образованию водородной связи в неполярной среде.

3.2. Аддитивно-мультипликативная модель энтальпии и свободной энергии водородной связи. Полифункциональные соединения

Представленные в разделе 3.1. результаты свидетельствуют о том, что аддитивно-мультипликативная модель термодинамических характеристик Н-комплексов и рассчитанные на ее основе дескрипторы Н-связи могут корректно представлять относительную протонодонорную и протоноакцепторную способность простых монофункциональных органических соединений. Однако, большинство физиологически активных веществ, лекарств представляют собой сложные соединения, в состав которых входит несколько функциональных групп. В литературе представлено недостаточно данных по термодинамическим характеристикам Н-комплексов с участием таких веществ. Поэтому с целью проверки работоспособности предлагаемой модели и дескрипторов в случае сложных органических соединений были проведены экспериментальные исследования. Они включали в себя установление стехиометрии, определение взаимодействующих атомов и групп, расчет величин ΔH , ΔG для Н-комплексов с участием фосфорилированных и тиофосфорилированных оксимонов, дифторнитроацетанилидов, фенола и ГМФТА.

В работе было использовано два ряда новых физиологически активных соединений. Один из них представлен 22 фосфорилированными и тиофосфорилированными оксимами (рис. 5, табл. 7), которые можно рассматривать в качестве аналогов инсектицида «валексон» [364]: $(EtO)_2P(S)ONC(C_6H_5)CN$.

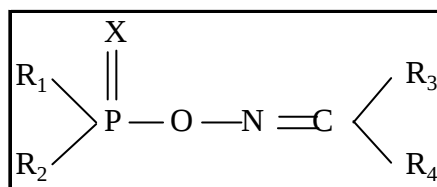


Рис. 5. Структурная формула фосфорилированных и тиофосфорилированных оксимов

Таблица 7

Номера соединений и заместители в фосфорилированных и тиофосфорилированных оксимах

№	R ₁	R ₂	R ₃	R ₄	X
I	Me	EtO	Me	Cl	O
II	Me	i-PrO	Me	Cl	O
III	Me	BuO	Me	Cl	O
IV	Me	AmO	Me	Cl	O
V	Me	EtO	Et	Cl	O
VI	Me	EtO	Pr	Cl	O
VII	Me	EtO	CH ₂ Cl	Cl	O
VIII	MeO	MeO	Me	Et	O
IX	Me	BuO	Et	Cl	O
X	Me	BuO	Pr	Cl	O
XI	MeO	MeO	Me	Me	S
XII	EtO	EtO	Me	Me	S
XIII	PrO	PrO	Me	Me	S
XIV	BuO	BuO	Me	Me	S
XV	MeO	MeO	Me	Et	S
XVI	EtO	EtO	Me	Et	S
XVII	PrO	PrO	Me	Et	S
XVIII	BuO	BuO	Me	Et	S
XIX	MeO	MeO	Me	Pr	S
XX	EtO	EtO	Me	Pr	S
XXI	PrO	PrO	Me	Pr	S
XXII	BuO	BuO	Me	Pr	S

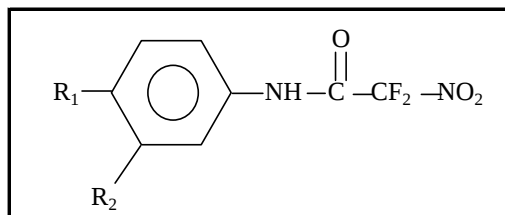


Рис. 6. Структурная формула дифторнитроацетанилидов

Таблица 8

Номера соединений и заместители в дифторнитроацетанилидах

№	R ₁	R ₂
XXIII	H	H
XXIV	H	CH ₃
XXV	H	F
XXVI	H	CF ₃
XXVII	H	NO ₂
XXVIII	CH ₃	H
XXIX	OCH ₃	H
XXX	F	H
XXXI	NO ₂	H
XXXII	Cl	CF ₃

В качестве другого ряда исследуемых соединений фигурировало 10 дифторнитроацетанилидов (рис. 6, табл. 8), которые можно рассматривать в качестве аналогов гербицида «пропанид»: $3,4\text{-Cl}_2\text{-C}_6\text{H}_3\text{-NH-C(O)-C}_2\text{H}_5$ [364].

При изучении ИК спектров растворов комплексов оксимов с фенолом в гексане наблюдается картина типичная при образовании водородной связи. Так, величина сдвига в низкочастотную область полосы валентных колебаний гидроксильной группы фенола ($\nu(\text{OH})=3624\text{ см}^{-1}$) составляет 197306 см^{-1} , а молярный коэффициент экстинкции этой полосы ($\epsilon=202\text{ л/моль см}$) возрастает более, чем в два раза. Также наблюдается сдвиг в низкочастотную область полосы валентных колебаний фосфорильной группы на $30\div 50\text{ см}^{-1}$, что свидетельствует об участии фосфорильной группы в Н-связывании.

При исследовании термодинамических характеристик комплексов (I)÷(XXII) с фенолом (1:1, гексан, 298 К) было установлено, что величины ΔH меняются от -18.5 до -28.2 кДж/моль, а величины ΔG от -6.1 до -16.3 кДж/моль. Это дало возможность вычислить величины дескрипторов E_a и C_a для исследованных соединений, которые варьировали в пределах от 1.19 до 1.83 и от 1.77 до 3.31 соответственно.

Дифторнитроацетанилиды (XXIII)÷(XXXII) содержат как протоноакцепторные (C=O), так и протонодонорные (N-H) группировки. При добавлении фенола к этим соединениям в CCl_4 или бензоле тепловые эффекты реакции малы. Однако из этого нельзя сделать вывод об отсутствии Н-акцепторных свойств у карбонильной группы, а можно лишь констатировать, что величины равновесных констант Н-связи соединений (XXIII)÷(XXXII) с фенолом незначительны и не превышают в указанных растворителях, по нашим оценкам, 3 л/моль.

Анализ протонодонорных свойств дифторнитроацетанилидов был проведен путем изучения их комплексообразования с ГМФТА в бензоле, работая в области таких концентраций Н-доноров, при которой их самоассоциацией можно было пренебречь. При этом наблюдаемый

низкочастотный сдвиг полос поглощения валентных колебаний $\nu(\text{NH})$ составлял от 103 до 128 см^{-1} . Интервал величин ΔH и ΔG комплексов (XXIII)÷(XXXII) с ГМФТА (1:1, бензол, 298 К) варьировал от -24.9 до -30.5 кДж/моль и от -15.0 до -20.8 кДж/моль соответственно. Рассчитанные величины E_d менялись в интервале -2.29÷-2.80, а величины C_d менялись в интервале -2.38÷-3.05.

В табл. 9 представлены результаты изучения взаимодействия по типу Н-связи между дифторнитроацетанилидами (XXV, XXVIII, XXXI) и фосфорилированными оксимами (III, IX, X) в бензоле при 298 К. Анализ термодинамических характеристик комплексообразования при взаимодействии одного Н-акцептора с несколькими Н-донорами или одного Н-донора с несколькими Н-акцепторами показывает, что в полученных рядах порядок расположения соединений по их протонодонорной или протоноакцепторной способности хорошо сохраняется (в случае ΔH эти изменения не выражены столь ярко из-за практической неизменности энтальпии изученных реакций). Такое поведение характерно в случае образования водородной связи с участием более простых модельных соединений.

При сравнении экспериментальных и рассчитанных на основе дескрипторов Н-связи термодинамических величин можно установить, что в случае ΔH минимальное отклонение (по абсолютной величине) составляет 0.2 кДж/моль, максимальное 1.3 кДж/моль и среднее 0.6 кДж/моль, а для ΔG 0.1 кДж/моль, 1.3 кДж/моль и 0.8 кДж/моль соответственно. Учитывая, что средние значения стандартных отклонений в серии измерений составили для ΔH и ΔG 1 кДж/моль и 0.6 кДж/моль соответственно, можно констатировать вполне удовлетворительное совпадение экспериментальных и рассчитанных термодинамических параметров комплексообразования.

Таблица 9

Экспериментальные и рассчитанные термодинамические характеристики Н-комплексов дифторнитроацетанилидов с фосфорилированными оксимами (бензол; 298 К; состав 1:1)

Н-донор	Н-акцептор	-ΔН, кДж/моль		-ΔG, кДж/моль	
		Эксп.	Расч.*	Эксп.	Расч.*
XXV	III	19.5	19.0	12.4	13.0
XXVIII	III	18.9	20.2	10.5	10.4
XXXI	III	18.9	18.5	13.8	14.4
XXV	IX	19.5	19.2	12.6	11.3
XXVIII	IX	19.4	20.4	9.98	9.24
XXXI	IX	18.9	18.7	13.5	12.4
XXV	X	19.6	19.4	13.3	12.5
XXVIII	X	19.8	20.6	11.0	10.3
XXXI	X	19.6	18.9	15.3	14.2

Примечание: * $\Delta H = 4.35 E_a E_d$; $\Delta G = 5.70 + 2.17 C_a C_d$; средние значения

стандартных отклонений в серии измерений составили для ΔН и ΔG 1 кДж/моль и 0.6 кДж/моль соответственно.

3.3. Модель липофильности органических соединений

Липофильность химических соединений рассматривается как важнейший физико-химический параметр практически во всех исследованиях, связанных с токсикологическими свойствами веществ и их фармакокинетическим поведением [365]. А логарифм коэффициента распределения соединений в системе н-октанол-вода ($\log K_{ow}$) является наиболее часто используемым дескриптором при установлении количественных соотношений «структура – активность» [306]. Отсюда очевидна необходимость разработки процедур расчета липофильности. Существует много подходов для расчета липофильности на основе: констант гидрофобности атомов [366, 367] или фрагментов [368, 369] (в том числе, конформационно зависимых констант [370]), графо-теоретических дескрипторов [371], структурных [372] и сольватохромных [373] параметров и большое количество компьютерных программ, основанных на использовании вышеуказанных подходов [374]. Следует отметить, что в качестве недостатков регрессионных моделей липофильности, использующих молекулярные дескрипторы, выступает большое количество параметров и плохая интерпретируемость независимых переменных.

В работах [375, 376] высказано предположение о том, что коэффициенты распределения вещества в системе н-октанол-вода отражают два главных структурных вклада: одного, связанного с молекулярным объемом, и другого, связанного с взаимодействием растворенного вещества с растворителем по типу водородной связи. В настоящее время существует достаточно много точных процедур и соответствующих компьютерных программ для количественной оценки параметров, связанных с молекулярным объемом. Что же касается количественного описания способности химических соединений образовывать водородные связи, ситуация до самого последнего времени являлась сложной и противоречивой. Здесь можно упомянуть о публикации

Сейлера с попыткой охарактеризовать вклад водородной связи в липофильность с помощью константы I_H [377], о работе Мorigучи и его параметре V_H [378], и о работах Абрахама и его сольватационных параметрах [365].

Как уже указывалось ранее, предложенные нами унифицированные шкалы энтальпийных и свободноэнергетических дескрипторов базируются на экспериментальных данных и рассчитываются на основе оценки термодинамических параметров с использованием уравнений (33) и (34). Проведенное нами сравнение экспериментальных и рассчитанных по уравнениям (33) и (34) значений энтальпии и свободной энергии Гиббса для трех тысяч Н-комплексов продемонстрировало приемлемость использования предложенного подхода для количественного описания термодинамики водородной связи:

$$\Delta H_{\text{расч}} = -0.49(\pm 0.29) + 0.99(\pm 0.08) \Delta H_{\text{эксп}} \quad (49)$$

$n=2787$; $r=0.970$; $s=2.40$; $F=44350$

$$\Delta G_{\text{расч}} = -0.07(\pm 0.12) + 1.04(\pm 0.05) \Delta G_{\text{эксп}} \quad (50)$$

$n=3301$; $r=0.991$; $s=1.12$; $F=175000$

Тем не менее, в ряде сильных Н-комплексов с участием атома азота подчас наблюдаются существенные отклонения рассчитанных по уравнению (34) значений свободной энергии Гиббса от экспериментальных данных. В качестве других недостатков предложенного подхода оценки относительной силы доноров и акцепторов водородной связи, здесь следует упомянуть о трудности идентификации атомов, непосредственно участвующих в образовании Н-связи, в случае сложных комплексов, о косвенном характере оценки силы слабых акцепторных центров, находящихся рядом с сильными, о невозможности

оценить акцепторную силу заряженных групп (поскольку уравнения (33) и (34) основаны на экспериментальных данных для неполярных растворов).

В связи с указанными недостатками описанной шкалы свободноэнергетических донорных и акцепторных дескрипторов для количественного описания липофильности были отобраны простые органические соединения, содержащие по одной акцепторной группе. Отобранные таким образом простые органические соединения, вместе с их значениями $\log K_{ow}$, молекулярной поляризуемости и акцепторной способности, представлены в табл. 10.

Таблица 10

Физико-химические дескрипторы: молекулярная поляризуемость A ($\text{\AA}^3$), протоноакцепторный свободноэнергетический дескриптор (C_a) и экспериментальные и рассчитанные значения липофильности ($\log K_{ow}$) соединений, содержащих одну функциональную группу

№	Соединение	α	C_a	$\log K_{ow}$	
				Расч.	Эксп.
1	Диметилсульфоксид	8.08	3.12	-0.97	-1.35
2	Метанол	3.25	1.76	-0.89	-0.77
3	Этанол	5.08	1.82	-0.47	-0.31
4	2-Пропанон	6.43	1.92	-0.21	-0.24
5	Оксетан	6.14	1.90	-0.26	-0.14
6	Метилоксиран	6.14	1.71	-0.07	0.03
7	2-Пропанол	6.92	1.78	0.06	0.05
8	1-Пропанол	7.76	1.60	0.47	0.25
9	2-Бутанон	8.27	1.93	0.27	0.29
10	2-Метил-2-пропанол	8.75	1.88	0.45	0.35
11	Тетрагидрофуран	7.98	2.01	0.12	0.46
12	Пропаналь	6.43	1.50	0.21	0.59
13	2-Бутанол	8.75	1.56	0.77	0.61
14	Пиррол	8.43	1.33	0.92	0.75
15	2-Метил-1-пропанол	8.75	1.75	0.58	0.76
16	Циклогексанон	11.16	2.03	0.94	0.81
17	3-Метил-2-бутанон	10.10	1.84	0.85	0.84
18	1-Бутанол	10.09	1.83	0.86	0.88
19	Бутаналь	9.11	1.58	0.85	0.88
20	Диэтиловый эфир	8.75	1.62	0.71	0.89
21	2-Пентанон	10.94	1.91	1.01	0.91
22	Метил-трет-бутиловый эфир	10.59	1.91	0.91	0.94
23	Тетрагидропиран	9.81	1.76	0.85	0.95
24	3-Пентанон	10.10	1.62	1.07	0.99
25	Этилен	4.25	0.20	0.93	1.13
26	3-Метил-1-бутанол	11.43	1.73	1.32	1.42

Продолжение табл. 10

№	Соединение	α	C_a	$\log K_{ow}$	
				Расч.	Эксп.
27	1-Метилпиррол	10.27	1.25	1.49	1.21
28	Циклогексанол	11.65	1.88	1.22	1.23
29	Тетрагидро-2,5-диметилфуран	11.65	2.11	0.99	1.28
30	4-Метил-2-пентанон	11.94	1.60	1.58	1.31
31	Фуран	7.23	0.64	1.29	1.34
32	2-Гексанон	13.28	2.03	1.51	1.38
33	Диизопропиловый эфир	12.42	1.78	1.53	1.52
34	Метилэтилсульфид	9.28	1.00	1.47	1.54
35	Метилбутиловый эфир	11.93	1.65	1.53	1.66
36	Пропилен	6.09	0.30	1.32	1.77
37	Тиофен	9.66	0.57	2.00	1.81
38	2-Хлор-2-метилпропан	10.04	0.67	2.01	1.82
39	2-Метилфуран	9.07	0.55	1.87	1.85
40	2,4-Диметил-3-пентанон	13.77	1.54	2.13	1.86
41	Диэтилсульфид	11.11	1.10	1.86	1.95
42	2-Гептанон	15.46	2.00	2.12	1.98
43	Дипропиловый эфир	13.26	1.66	1.87	2.03
44	Этилбутиловый эфир	13.76	1.84	1.83	2.03
45	2,5-Диметилфуран	10.90	0.85	2.05	2.24
46	2-Метилтиофен	11.49	0.60	2.46	2.33
47	2-Хлорбутан	10.04	0.54	2.14	2.33
48	1-Фторпентан	11.55	0.75	2.33	2.33
49	2-Октанон	17.57	2.00	2.68	2.37
50	1-Хлорбутан	11.38	0.60	2.43	2.64
51	Метилбензол	12.27	0.48	2.79	2.73
52	1-Гексин	12.02	0.60	2.60	2.73
53	1-Бромбутан	12.08	0.50	2.72	2.75
54	Циклогексен	10.82	0.40	2.48	2.86
55	1-Октанол	18.46	1.79	3.13	3.00
56	2,2,4,4-Тетраметил-3-пентанон	17.44	1.78	2.87	3.00
57	1,2-Диметилбензол	14.10	0.72	3.04	3.12
58	Этилбензол	14.10	0.54	3.22	3.15

Продолжение табл. 10

№	Соединение	α	C_a	$\log K_{ow}$	
				Расч.	Эксп.
59	1,4-Диметилбензол	14.10	0.60	3.16	3.15
60	Бромциклогексан	13.64	0.58	3.05	3.20
61	1,3-Диметилбензол	14.10	0.54	3.22	3.20
62	Дибутиловый эфир	17.43	1.48	3.16	3.21
63	Циклопропилбензол	15.17	0.54	3.50	3.27
64	1-Бромпентан	14.27	0.75	3.05	3.37
65	1,3,5-Триметилбензол	15.94	0.66	3.59	3.42
66	Изопропилбензол	15.94	0.54	3.71	3.66
67	1-Бромгексан	16.37	0.51	3.85	3.80
68	1,2,4,5-Тетраметилбензол	17.77	0.84	3.89	4.00
69	Трет-бутилбензол	17.77	0.60	4.13	4.11
70	1-Бромгептан	18.43	0.52	4.39	4.36
71	Гексаметилбензол	21.44	0.96	4.75	4.61

Результирующее двухпараметровое уравнение, где в качестве зависимой переменной выступает липофильность, а в качестве независимых переменных – поляризуемость и протонноакцепторная способность химических соединений, представлено ниже:

$$\log K_{ow} = 0.266(\pm 0.006) \alpha - 1.00(\pm 0.05) C_a \quad (51)$$

$n=71; r^2=0.982; s=0.18; F=3829; q^2=0.981$

Статистические критерии уравнения (51) свидетельствуют о возможности количественно оценки значений $\log K_{ow}$ на основе рассчитанных величин α и C_a .

Для того, чтобы проверить устойчивость полученной модели, была проведена ее валидация путем расчета аналогичной модели с использованием другого ряда соединений. В табл. 11 приведены характеристики этого ряда.

Таблица 11

Значения коэффициента распределения вещества в системе н-октанол-вода ($\log K_{ow}$), молекулярной поляризуемости, $\bar{A}^{(3)}$ и протоноакцепторного свободноэнергетического дескриптора (C_a) органических соединений

№	Соединение	α	C_a	$\log K_{ow}$
1	1,1,1-Трихлорэтан	10.23	0.21	2.49
2	1,1,2-Трихлорэтан	10.23	0.68	1.89
3	1,1-Дихлорэтан	8.30	0.31	1.79
4	1,2,3-Трихлорбензол	16.22	0.24	4.14
5	1,2,4,5-Тетрахлорбензол	18.15	0.23	4.60
6	1,2-Дихлорбензол	14.29	0.37	3.43
7	1,2-Дихлорэтан	8.30	0.78	1.48
8	1,3,5-Трихлорбензол	16.22	0.24	4.19
9	1,4-Диоксан	8.61	2.96	-0.27
10	1,4-Дихлорбензол	14.29	0.37	3.44
11	1-Бутанол	8.75	1.60	0.88
12	1-Гексанол	12.42	1.59	2.03
13	1-Гексен	11.59	0.20	3.39
14	1-Нонен	17.10	0.20	5.15
15	1-Октанол	16.09	1.61	3.00
16	1-Октен	15.26	0.20	4.57
17	1-Пентанол	10.59	1.61	1.56
18	1-Пропанол	6.92	1.61	0.25
19	1-Хлорбутан	10.04	0.39	2.64
20	2-Бутанол	8.75	1.62	0.61
21	2-Бутанон	8.27	1.91	0.29
22	2-Гексанон	11.94	1.90	1.38
23	2-Гептанон	13.77	1.93	1.98
24	2-Метил-1-пропанол	8.75	1.57	0.76
25	2-Метил-2-пропанол	8.75	1.60	0.35
26	2-Пентанон	10.10	1.90	0.91
27	2-Пропанол	6.92	1.69	0.05
28	3-Метил-1-бутанол	10.59	1.62	1.16

Продолжение табл. 11

№	Соединение	α	C_a	$\log K_{ow}$
29	4-Нитроанилин	13.75	2.39	1.39
30	cis-1,2-Дихлорэтен	8.11	0.40	1.86
31	N,N-Диметилформамид	7.91	2.87	-1.01
32	Анизол	12.54	1.19	2.11
33	Аценафтен	20.60	1.56	3.92
34	Ацетон	6.43	1.95	-0.24
35	Ацетонитрил	4.47	1.73	-0.34
36	Бензол	10.43	0.64	2.13
37	Бензонитрил	12.29	1.84	1.56
38	Бифенил	20.09	1.32	4.01
39	Бутан	8.11	0.00	2.89
40	Бутилацетат	12.21	1.62	1.78
41	Бутиральдегид	8.27	1.54	0.88
42	Бутиронитрил	8.14	1.66	1.12
43	Гексаналь	11.94	1.64	1.78
44	Дибромметан	7.86	0.30	1.88
45	Дибутиловый эфир	16.09	1.53	3.21
46	Диизопропиловый эфир	12.42	1.58	1.52
47	Диметиламин	5.80	1.92	-0.38
48	Диметиловый эфир	5.08	1.45	0.10
49	Дипропиламин	13.14	2.05	1.67
50	Дипропиловый эфир	12.42	1.50	2.03
51	Дихлорметан	6.47	0.30	1.25
52	Диэтиловый эфир	8.75	1.54	0.89
53	Изопропилбензол	15.94	0.68	3.66
54	Метан	2.61	0.00	1.09
55	Метанол	3.25	1.63	-0.77
56	Метилацетат	6.71	1.65	0.18
57	Метилиодид	7.64	0.32	1.51
58	Метил-трет-бутиловый эфир	10.59	1.58	0.94
59	Метилформиат	4.87	1.51	0.03
60	м-Ксилол	14.10	0.45	3.20
61	Нафталин	17.70	1.40	3.30

Продолжение табл. 11

№	Соединение	α	C_a	$\log K_{ow}$
62	н-Бутиламин	9.47	2.00	0.97
63	Нитрометан	4.32	1.40	-0.35
64	Нитропропан	7.99	1.40	0.87
65	Нитроэтан	6.16	1.40	0.18
66	н-Пропилхлорид	8.21	0.37	2.04
67	о-Ксилол	14.10	0.63	3.12
68	Пентан	9.95	0.00	3.39
69	Пирен	28.77	2.24	4.88
70	п-Ксилол	14.10	0.58	3.15
71	Пропан	6.28	0.00	2.36
72	Пропаналь	6.43	1.52	0.59
73	Пропилацетат	10.38	1.65	1.24
74	Пропилбензол	15.94	0.74	3.72
75	Пропилформиат	8.54	1.54	0.83
76	Пропионитрил	6.30	1.72	0.53
77	Тетрагидропиран	9.81	1.66	0.95
78	Тетрагидрофуран	7.98	1.66	0.46
79	Тетрахлорэтилен	11.96	0.23	3.40
80	Толуол	12.27	0.48	2.73
81	Уксусная кислота	4.87	1.51	-0.17
82	Фенантрен	24.97	1.96	4.46
83	Флуорен	21.16	1.29	4.18
84	Хлорметан	4.54	0.40	0.91
85	Хлороформ	8.39	0.24	1.97
86	Циклогексан	11.01	0.00	3.44
87	Циклогексано́л	11.65	1.67	1.23
88	Циклогексен	10.82	0.20	2.86
89	Четырехфтористый углерод	2.25	0.39	1.18
90	Четыреххлористый углерод	10.32	0.64	2.83
91	Этан	4.44	0.00	1.90
92	Этанол	5.08	1.66	-0.31
93	Этен	4.25	0.20	1.13
94	Этилацетат	8.54	1.69	0.73

Продолжение табл. 11

№	Соединение	α	C_a	$\log K_{ow}$
95	Этилбензол	14.10	0.61	3.15
96	Этилбромид	7.07	0.40	1.61
97	Этилиодид	9.47	0.32	2.00
98	Этилпропаноат	10.38	1.65	1.21

В результате было получено следующее уравнение:

$$\log K_{ow} = 0.276(\pm 0.008) \alpha - 1.01(\pm 0.07) C_a \quad (52)$$

$$n=98; r^2=0.961; s=0.29; F=1191; q^2=0.959$$

Параметры уравнений (51) и (52) близки, что свидетельствует об устойчивости модели (51).

Установленная зависимость (уравнение (51)) позволяет также оценивать вклады поляризуемости и Н-акцепторной способности в липофильность соединений. Так, отрицательные экспериментальные значения $\log K_{ow}$ для первых пяти соединений в табл. 10 свидетельствуют о превалировании вклада протоноакцепторной способности (имеющего всегда отрицательное значение) над вкладом поляризуемости (имеющего всегда положительное значение), а положительные значения $\log K_{ow}$ для соединений № 8-71 в табл. 10 свидетельствуют о преобладании вклада поляризуемости над вкладом Н-акцепторной способности.

Преобразование уравнения (51) к следующему виду:

$$C_{a(o)} = 0.266 \alpha - \log K_{ow} \quad (53)$$

дает возможность построить новую шкалу свободноэнергетических протоноакцепторных дескрипторов ($C_{a(o)}$), основанную на экспериментальных данных по липофильности. Для простых соединений, содержащих одну

акцепторную группу, значения C_a и $C_{a(o)}$ будут одинаковы. В случае же сложных органических соединений, $C_{a(o)}$ будет соответствовать сумме свободноэнергетических протоноакцепторных дескрипторов, а C_a – значению дескриптора для наиболее сильного Н-акцептора в молекуле.

Имеется много опубликованных данных по экспериментальным величинам $\log K_{ow}$ как простых, так и сложных нейтральных и ионизированных соединений. И это дает возможность на основе уравнения (53) оценить протоноакцепторную способность тех атомов и функциональных групп, которые не могли быть определены в рамках шкалы, основанной на термодинамических параметрах. Так, в табл. 12 приведены оцененные значения C_a^* для слабых Н-акцепторов, находящихся по соседству с сильными Н-акцепторами. Как видно из данных этой таблицы, такие слабые акцепторы вносят небольшой, но заметный вклад в общую протоноакцепторную способность соединений и, следовательно, в их липофильность. Подобная работа была проведена для нескольких тысяч органических соединений. В результате удалось существенно увеличить число соединений, для которых была количественно охарактеризована протоноакцепторная способность к образованию Н-связи.

Таблица 12

Суммарные величины свободноэнергетических дескрипторов ($C_{a(o)}$), дескрипторы сильных (C_a) и слабых (C_a^*) протоноакцепторов для органических соединений, содержащих функциональные группы с несколькими гетероатомами

№	Соединение	$C_{a(o)}$	C_a	C_a^*
1	м-Нитробензойная кислота	0.62	0.47	0.15
2	Бензойная кислота	1.03	0.76	0.27
3	Этилтрихлорацетат	1.21	1.11	0.10
4	Уксусная кислота	1.47	1.10	0.37
5	Метилацетат	1.60	1.40	0.20
6	Бутиролактон	2.71	2.05	0.66
7	N,N-Диметилацетамид	3.02	2.76	0.26
8	Мочевина	3.06	2.50	0.56
9	N-Метилацетамид	3.15	2.70	0.45
10	2-Пиперидон	3.34	2.95	0.39

3.4. Компьютерный программный комплекс НУВОТ

НУВОТ предназначен для количественной оценки Н-донорной и Н-акцепторной способности органических соединений путем расчета дескрипторов водородной связи. Этот программный комплекс состоит из трех главных компонентов. Во-первых, это база данных по термодинамике водородной связи, которая содержит экспериментальные величины энтальпии и свободной энергии Гиббса образования Н-комплексов в различных средах (рис. 7). В настоящее время эта БД насчитывает 13688 записей (т.е. сведений об одной паре реакций Н-донор/Н-акцептор).

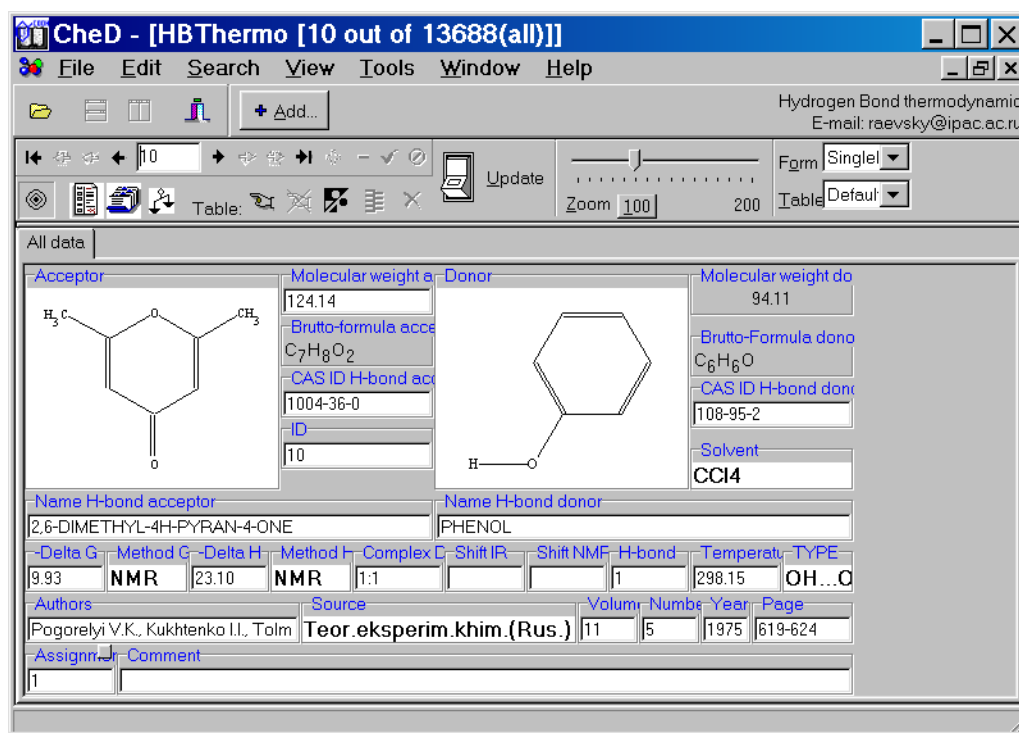


Рис. 7. Пример представления информации в термодинамической БД для одной записи

БД содержит ряд полей (табл. 13).

Таблица 13

Поля в термодинамической БД

Обозначение	Название
-Delta G	Свободная энергия Гиббса образования Н-комплекса, кДж/моль
-Delta H	Энтальпия образования Н-комплекса, кДж/моль
Method G	Метод определения свободной энергии Гиббса
Method H	Метод определения энтальпии
Shift NMR	Разница в химических сдвигах между свободными и связанными Н-донорными группами (ppm)
Shift IR	Разница между частотами колебаний свободных и связанных Н-донорных групп (см ⁻¹)
Type	Тип водородного мостика в комплексе АН...В, где АН – протонодонорная группа, В – протоноакцепторный атом
Temperature	Температура, к которой относятся ΔН и ΔG
Solvent	Растворитель
Complex D/A	Стехиометрия Н-комплекса
H-bond	Число водородных мостиков в Н-комплексе
Comment	Комментарии к ссылке и данным
Authors	Авторы публикации
Source	Наименование источника данных
Year	Год публикации
Volume	Том журнала
Number	Номер журнала
Page	Страницы публикации
Molecular weight	Молекулярная масса соединения
Name	Название химического соединения
CAS ID	Номер CAS

Продолжение табл. 13

Обозначение	Название
Brutto formula	Брутто-формула соединения
Donor	Химическая структура Н-донора
Acceptor	Химическая структура Н-акцептора

Другой составной частью НУВОТ является дескрипторная база данных, содержащая протонодонорные и протоноакцепторные дескрипторы органических соединений, рассчитанных в рамках общей шкалы водородной связи. Эта БД содержит 4 библиотеки дескрипторов: E_d (1025 дескрипторов), E_a (2060 дескрипторов), C_d (2835 дескрипторов), C_a (46372 дескриптора) (рис. 8).

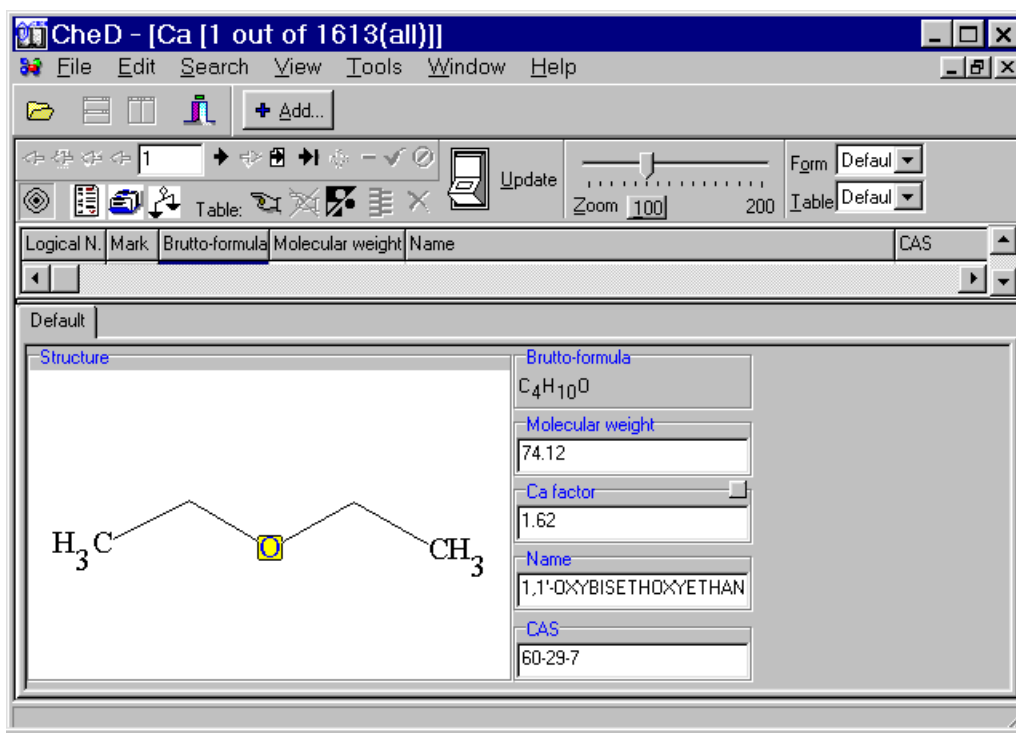


Рис. 8. Пример представления информации в дескрипторной БД для одного дескриптора

БД содержит ряд полей (табл. 14).

Таблица 14

Поля в дескрипторной БД

Обозначение	Название
Structure	Химическая структура соединения
Name	Название соединения
Brutto formula	Брутто-формула соединения
Molecular weight	Молекулярная масса соединения
CAS	Номер CAS
Ca factor	Величина дескриптора

Третьей составной частью HUBOT является программа, которая рассчитывает дескрипторы органических соединений на основе их химической структуры. Величина дескриптора оценивается с использованием анализа химического структурного окружения [379]. Суть этого подхода состоит в генерации структурных фрагментов, состоящих из центрального атома и соседних атомов, связанных с ним в пределах сферы определенного размера (число связей между центральным атомом и конечными атомами). В этих фрагментах каждый атом характеризуется тремя параметрами: парциальным атомным зарядом, поляризуемостью и Н-донорным/акцепторным дескриптором. В случае полного совпадения структуры интересующей нас молекулы со структурой из дескрипторной БД соответствующее значение дескриптора приписывается исследуемой молекуле. В противном случае проводится статистическая оценка дескрипторов подобных структур. Всего может быть рассчитано 32 дескриптора Н-связи (рис. 9, табл. 15).

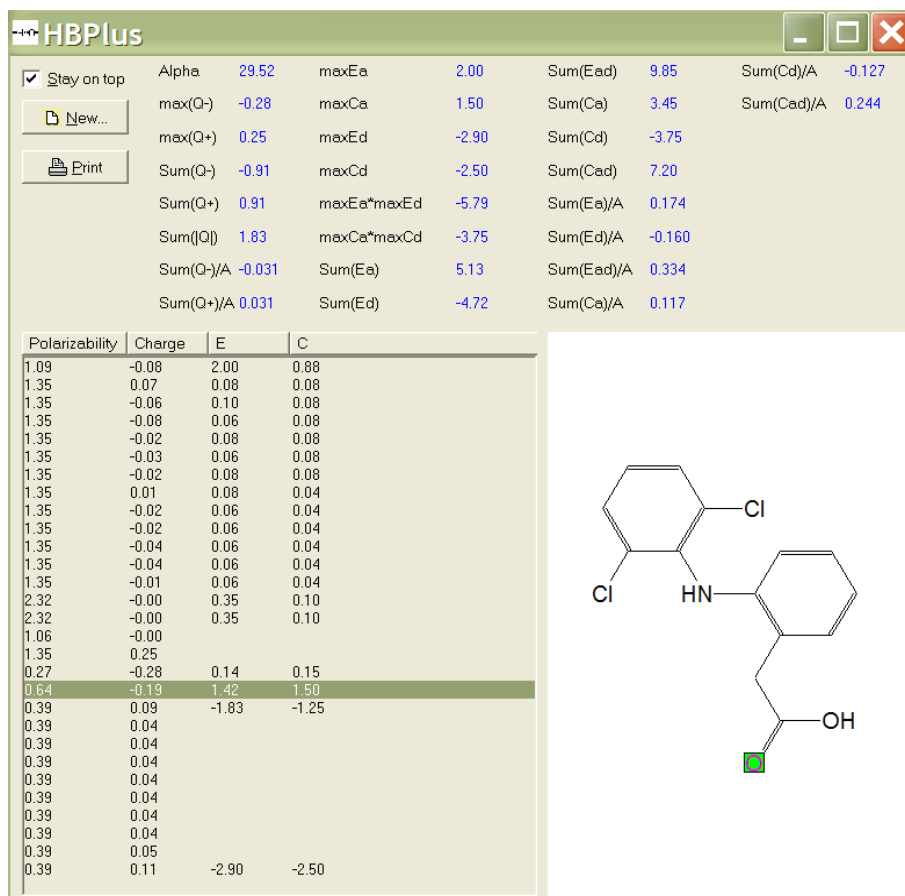


Рис. 9. Дескрипторы Н-связи для молекулы диклофенака

Таблица 15

Дескрипторы HUBOT

Обозначение	Дескриптор
Alpha	Молекулярная поляризуемость, $\text{\AA}^3$
max(Ca)	Максимальный дескриптор C_a
max(Cd)	Максимальный дескриптор C_d
max(Q+)	Максимальный положительный атомный заряд
max(Q-)	Максимальный отрицательный атомный заряд
Sum(Ca)	Сумма дескрипторов C_a
Sum(Cd)	Сумма дескрипторов C_d
Sum(Q+)	Сумма положительных атомных зарядов
Sum(Q-)	Сумма отрицательных атомных зарядов
Sum(Abs(Q))	Сумма абсолютных значений атомных зарядов
Sum(Ea)	Сумма дескрипторов E_a
Sum(Q+)/A	Сумма положительных атомных зарядов/Alpha
Sum(Q-)/A	Сумма отрицательных атомных зарядов/Alpha
Sum(Ca)/A	Сумма дескрипторов C_a /Alpha
Sum(Cd)/A	Сумма дескрипторов C_d /Alpha
Sum(Ed)	Сумма дескрипторов E_d

Продолжение табл. 15

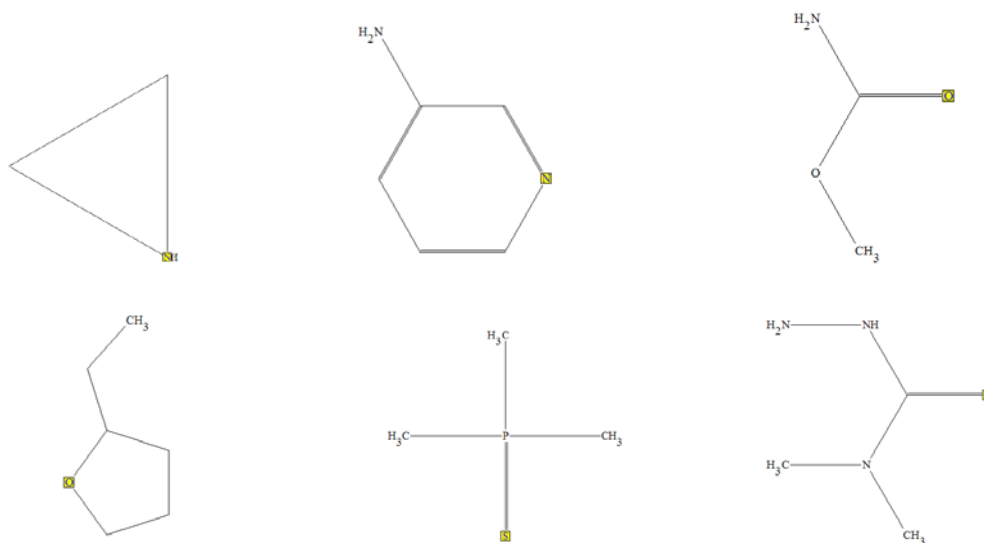
Обозначение	Дескриптор
Max(Ea)	Максимальный дескриптор E _a
Max(Ed)	Максимальный дескриптор E _d
max(Ca(o))	Максимальный дескриптор C _a (o)
MaxEa*maxEd	Произведение максимальных дескрипторов E _a E _d
MaxCa*maxCd	Произведение максимальных дескрипторов C _a C _d
MaxCa(o)*maxCd	Произведение максимальных дескрипторов C _a (o)C _d
Sum(Ead)	Сумма дескрипторов E
Sum(Cad)	Сумма дескрипторов C
Sum(Ca(o))	Сумма дескрипторов C _a (o)
Sum(Cad(o))	Сумма дескрипторов C(o)
Sum(Ea)/A	Сумма дескрипторов E _a /Alpha
Sum(Ed)/A	Сумма дескрипторов E _d /Alpha
Sum(Ead)/A	Сумма дескрипторов E/Alpha
Sum(Cad)/A	Сумма дескрипторов C/Alpha
Sum(Ca(o))/A	Сумма дескрипторов C _a (o)/Alpha
Sum(Cad(o))/A	Сумма дескрипторов C(o)/Alpha

Так как водородная связь играет важную роль для проявления многих видов физико-химических свойств и биологической активности, ее учет при проектировании соединений для целевого органического синтеза является необходимой и важной задачей. Дизайн новых органических соединений с использованием HUBOT может быть проведен как путем решения прямой задачи (расчет дескрипторов Н-связи для ряда гипотетических структур), так и путем решения обратной задачи (генерация структур/подструктур с заданной величиной Н-дескриптора). Пример расчета молекулы диклофенака в рамках прямой задачи приведен на рис. 9.

Из представленных данных, в частности, следует, что максимальные величины дескрипторов C_a, E_a, C_d, E_d составляют 1.50, 2.00, -2.50, -2.90 соответственно. При этом произведена полная идентификация Н-активных центров (атомов) в молекуле. Так выделенный атом кислорода является Н-акцепторным центром и характеризуется величинами C_a=1.42 и E_a=1.50.

В рамках решения обратной задачи рассмотрим случай, когда для органического синтеза необходимо иметь набор подструктур с заданной

величиной Н-акцепторной активности $C_a=2.00$. С помощью НУВОТ было найдено 375 соответствующих структур. Некоторые из них приведены ниже.



Из представленных данных очевидно, что одинаковую величину Н-акцепторной активности могут иметь различные по природе атомы, с различным типом гибридизации и молекулярным окружением, входящие в состав ациклических, циклических и ароматических структур.

Выводы по главе

Разработана аддитивно-мультипликативная модель водородной связи с использованием величин энтальпии и свободной энергии 936 Н-комплексов состава 1:1 с одним водородным мостиком в CCl_4 и рассчитаны значения дескрипторов водородной связи, характеризующих способность монофункциональных органических соединений к образованию Н-связи.

Выявлено, что способность соединений к образованию Н-связи необходимо описывать с использованием двух типов дескрипторов: энтальпийных и свободноэнергетических.

Исследованы спектроскопические и термодинамические характеристики Н-комплексов фосфорилированных и тиофосфорилированных оксимов с фенолом, дифторнитроацетанилидов с ГМФТА, фосфорилированных оксимов с дифторнитроацетанилидами и установлено, что дескрипторы Н-связи могут

быть использованы для описания способности этих физиологически активных полифункциональных соединений к образованию водородной связи.

Разработана регрессионная QSPR модель липофильности 169 органических соединений, связывающая константы распределения вещества в системе н-октанол – вода с молекулярной поляризуемостью и свободноэнергетическими протоноакцепторными дескрипторами.

Создана база данных по термодинамическим характеристикам образования Н-комплексов, насчитывающая свыше 13000 записей.

Сформирована база данных по протонодонорным и протоноакцепторным дескрипторам Н-связи, содержащая свыше 50000 записей, и на ее основе создана компьютерная программа HYBOT, предназначенная для расчета дескрипторов водородной связи органических соединений.

Глава 4. Количественные локальные модели «структура – биологическая активность» органических соединений

4.1. Классификационные и регрессионные модели неполярного наркоза

Около половины известных индустриальных поллютантов вызывают у водных организмов острые токсические эффекты, соответствующие концепции «базовой токсичности» [380] или «неполярной неспецифической токсичности» [276]. Исследования QSAR показывают, что такая токсичность обычно хорошо коррелирует с липофильностью, выражаемой через $\log K_{ow}$. Другая часть поллютантов обладает более высокой токсичностью, такой как полярная неспецифическая токсичность, разобщение оксидативного фосфорилирования, ингибирование АХЭ, ингибирование дыхания, SH алкилирование, реактивная токсичность и эстрогенная активность [276]. Токсичность соединений может быть также ниже базовой, например, благодаря их метаболизму. Ясно, что QSAR модели для каждой из этих MOA, вероятно, должны быть основаны на различных рядах дескрипторов и различных типах уравнений. Таким образом, сконструировать общие регрессионные модели острой токсичности для разнообразных химических соединений, обладающих различными типами действия, не представляется возможным. В качестве альтернативы можно рассматривать создание классификационных моделей для распознавания MOA с последующей разработкой регрессионных моделей для соединений с одинаковым типом токсического действия.

Табл. 16 содержит данные по острой токсичности (LC_{50} , ммоль/л) 118 органических соединений с MOA неполярный наркоз по отношению к рыбам *Poecilia reticulata* (Pr), *Pimephales promelas* (Pp) и *Oncorhynchus mykiss* (Om), а также экспериментальные величины $\log K_{ow}$ и три физико-химических дескриптора: α ($\text{\AA}^3$), $\sum C_a$, $\sum C_d$.

Таблица 16

Острая токсичность (LC_{50} , ммоль/л), липофильность ($\log K_{ow}$), поляризуемость (α , Å³), сумма свободноэнергетических Н-акцепторных (ΣC_a) и Н-донорных (ΣC_d) дескрипторов органических соединений с МОА неполярный наркоз по отношению к *P. reticulata* (Pr), *P. promelas* (Pp) и *O. mykiss* (Om)

№	Соединение	CAS	$\log K_{ow}$	α	ΣC_a	ΣC_d	$\log(1/LC_{50})$		
							Pr	Pp	Om
1	Этилбензол	100-41-4	3.15	14.10	0.61	0.00	-	-	1.40
2	2-Этилпиридин	100-71-0	1.69	13.40	2.10	0.00	-	-0.59	-
3	Дифениловый эфир	101-84-8	4.21	20.37	1.39	0.00	1.62	-	-
4	4-Ксилол	106-42-3	3.15	14.10	0.58	0.00	0.48	-	1.61
5	4-Хлортолуол	106-43-4	3.33	14.20	0.60	0.00	1.33	-	-
6	1,4-Дихлорбензол	106-46-7	3.44	14.29	0.37	0.00	1.56	-	2.12
7	1-Бромпропан	106-94-5	2.10	8.90	0.49	0.00	-	0.26	-
8	1,2-Дихлорэтан	107-06-2	1.48	8.30	0.78	0.00	-0.06	-0.14	0.64
9	Пропионитрил	107-12-0	0.16	6.30	1.72	0.00	-	-1.44	-
10	1,2-Этандиол	107-21-1	-1.36	5.72	3.12	-3.16	-2.90	-	-
11	2-Метил-2,4-пентандиол	107-41-5	-0.67	13.06	3.32	-3.00	-1.96	-1.96	-1.90
12	2-Пентанон	107-87-9	0.91	10.10	1.90	0.00	-	-1.16	-
13	4-Метил-2-пентанон	108-10-1	1.31	11.94	1.87	0.00	-0.71	-0.72	-
14	Диизопропиловый эфир	108-20-3	1.52	12.42	1.58	0.00	0.04	-	-
15	3-Ксилол	108-38-3	3.20	14.10	0.45	0.00	0.45	-	1.10
16	3-Хлортолуол	108-41-8	3.28	14.20	0.54	0.00	0.84	-	-
17	1,3,5-Трихлорбензол	108-70-3	4.18	16.22	0.24	0.00	1.74	-	-
18	Толуол	108-88-3	2.78	12.27	0.48	0.00	0.13	-	1.20
19	Хлорбензол	108-90-7	2.89	12.36	0.53	0.00	0.77	0.82	1.38
20	Циклогексанол	108-93-0	1.23	11.65	1.67	-1.52	-0.85	-0.85	-
21	Циклогексанон	108-94-1	0.81	11.16	2.06	0.00	-0.73	-	-
22	2-Пиколин	109-06-8	1.11	11.56	2.19	0.00	-	-0.98	-
23	2-Изопропоксиэтанол	109-59-1	0.05	11.22	3.10	-1.63	-1.72	-	-
24	1-Бромбутан	109-65-9	2.75	10.74	0.46	0.00	-	0.57	-
25	1-Хлорбутан	109-69-3	2.64	10.04	0.39	0.00	-0.02	-	-
26	2-Метоксиэтанол	109-86-4	-0.77	7.55	3.10	-1.57	-2.36	-	-
27	Тетрагидрофуран	109-99-9	0.46	7.98	1.66	0.00	-1.48	-	-
28	Фуран	110-00-9	1.34	7.23	0.58	0.00	0.04	-	-
29	2-Гептанон	110-43-0	1.98	13.77	1.93	0.00	-	-0.06	-
30	2-Этоксиэтанол	110-80-5	-0.10	9.39	3.10	-1.58	-2.26	-	-
31	6-Метил-5-гептен-2-он	110-93-0	1.82	15.42	2.10	0.00	0.16	-	-

Продолжение табл. 16

№	Соединение	CAS	logK _{ow}	α	ΣC_a	ΣC_d	log(1/LC ₅₀)		
							Pr	Pp	Om
32	2-Октанон	111-13-7	2.37	15.61	1.95	0.00	0.55	0.55	-
33	1-Бромгексан	111-25-1	3.80	14.41	0.40	0.00	-	1.68	-
34	1-Гексанол	111-27-3	2.03	12.42	1.59	-1.59	0.02	0.02	-
35	Диэтиленгликоль	111-46-6	-1.30	13.70	4.76	-3.02	-2.76	-	-
36	2-Бутоксиэтанол	111-76-2	0.83	13.06	2.97	-1.58	-0.92	-	-
37	1-Бромооктан	111-83-1	4.89	18.08	0.45	0.00	-	2.36	-
38	1-Октанол	111-87-5	3.00	16.09	1.61	-1.35	0.98	0.98	0.87
39	2-(2-Этоксипропан-2-ил)этанол	111-90-0	-0.54	13.70	4.35	-1.86	-2.30	-	-2.00
40	2-Ундеканон	112-12-9	4.09	21.11	1.90	0.00	-	2.06	1.75
41	Триэтиленгликоль	112-27-6	-1.24	14.33	5.77	-3.16	-2.65	-	-
42	1-Деканол	112-30-1	4.57	19.76	1.59	-1.43	1.81	1.82	1.61
43	Бутилдигол	112-34-5	0.56	17.36	4.46	-1.58	-0.85	-	-
44	1-Ундеканол	112-42-5	4.53	21.60	1.59	-1.43	2.21	-	-
45	1-Додеканол	112-53-8	5.13	23.43	1.59	-1.43	2.26	-	-
46	2-Метил-3-Бутин-2-ол	115-19-5	0.28	9.48	1.94	-1.87	-	-1.59	-
47	2,2,2-Трихлорэтанол	115-20-8	1.42	10.87	1.62	-2.14	-0.31	-0.30	-0.19
48	Бензофенон	119-61-9	3.18	22.08	2.26	0.00	1.07	-	-
49	1,2,4-Трихлорбензол	120-82-1	4.02	16.22	0.28	0.00	1.83	1.78	2.07
50	N,N-Диметиланилин	121-69-7	2.31	15.58	1.88	0.00	-	0.28	-
51	Дифениламин	122-39-4	3.50	21.57	1.96	-1.26	-	1.65	-
52	2-Феноксипропанол	122-99-6	1.16	15.02	2.86	-1.65	-0.40	-0.40	-
53	Тетрахлорэтен	127-18-4	3.40	11.96	0.23	0.00	1.02	1.00	1.45
54	2,3,4-Триметоксиацетофенон	13909-73-4	1.63	20.59	4.01	0.00	0.08	0.02	-
55	1,3-Дихлорпропан	142-28-9	2.00	10.14	0.86	0.00	0.13	0.01	-
56	Дибутиловый эфир	142-96-1	3.21	16.09	1.53	0.00	0.60	-	-
57	1-Нонанол	143-08-8	4.26	17.93	1.59	-1.43	1.40	1.40	-
58	2,2,5,5-Тетраметилтетрагидрофуран	15045-43-9	2.06	15.32	1.62	0.00	-	-0.12	-
59	1,4-Диметоксибензол	150-78-7	2.03	14.65	1.89	0.00	0.07	-	-
60	трет-Бутилметиловый эфир	1634-04-4	0.94	10.59	1.58	0.00	-0.91	-	-
61	2,6-Дихлорбензамид	2008-58-4	0.77	17.76	3.13	-4.16	-	-0.39	-0.09
62	2,4-Дихлорацетофенон	2234-16-4	2.73	18.11	2.07	0.00	1.20	1.21	-
63	2,6-Диизопропиланилин	24544-04-5	3.18	22.92	1.94	-2.57	-	1.06	-
64	3-Фуранметанол	4412-91-3	0.30	9.70	2.08	-1.59	-0.72	-	-
65	2,3-Дигидробензофуран	496-16-2	2.14	13.60	1.29	0.00	-	0.17	-
66	5-Нонанон	502-56-7	2.91	17.44	1.97	0.00	0.66	-	-
67	Уретан	51-79-6	-0.15	8.18	2.44	-2.57	-	-1.77	-
68	1,3-Дихлорбензол	541-73-1	3.53	14.29	0.32	0.00	1.28	1.26	-
69	2-Гидрокси-4-метоксиацетофенон	552-41-0	1.98	16.64	2.52	-2.11	0.48	-	-

Продолжение табл. 16

№	Соединение	CAS	logK _{ow}	α	ΣC_a	ΣC_d	log(1/LC ₅₀)		
							Pr	Pp	Om
70	Тетрахлорметан	56-23-5	2.83	10.32	0.64	0.00	0.36	-	-
71	3-Метил-2-бутанон	563-80-4	0.84	10.10	1.88	0.00	-1.01	-	-
72	2,6-Диметокситолуол	5673-07-4	2.87	16.49	1.66	0.00	0.87	0.88	-
73	3-Пентанол	584-02-1	1.21	10.59	1.58	-1.37	-1.05	-	-
74	2-Гексанон	591-78-6	1.38	11.94	1.90	0.00	-	-0.63	-
75	2,4-Гексадиен	592-46-1	3.01	11.40	0.22	0.00	-	0.61	-
76	Диэтиловый эфир	60-29-7	0.89	8.75	1.54	0.00	-1.46	-1.54	-
77	Пентахлорбензол	608-93-5	5.18	20.07	0.26	0.00	3.15	-	2.97
78	2,5-Диметилфуран	625-86-5	2.24	10.90	0.66	0.00	-	0.13	-
79	1-Бромгептан	629-04-9	4.36	16.25	0.44	0.00	-	2.09	-
80	1,2,3,4-Тетрахлорбензол	634-66-2	4.64	18.15	0.25	0.00	2.35	-	-
81	1,2,3,5-Тетрахлорбензол	634-90-2	4.65	18.15	0.18	0.00	2.43	-	-
82	Этанол	64-17-5	-0.31	5.08	1.66	-1.42	-2.44	-2.50	-2.35
83	2,4,5-Трихлортолуол	6639-30-1	4.78	18.05	0.40	0.00	2.06	-	-
84	Метанол	67-56-1	-0.77	3.25	1.63	-1.50	-2.94	-2.96	-2.80
85	2-Пропанол	67-63-0	0.05	6.92	1.69	-1.25	-2.16	-2.16	-
86	2-Пропанон	67-64-1	-0.24	6.43	1.95	0.00	-2.10	-	-2.43
87	Хлороформ	67-66-3	1.97	8.39	0.24	0.00	0.07	0.23	0.73
88	Гексахлорэтан	67-72-1	4.14	16.01	0.30	0.00	2.19	2.22	2.45
89	2-Деканон	693-54-9	3.73	19.28	1.93	0.00	1.43	-	-
90	Дипентиловый эфир	693-65-2	4.04	19.76	1.53	0.00	1.69	-	-
91	1-Бутанол	71-36-3	0.88	8.75	1.60	-1.32	-1.37	-1.37	-
92	Бензол	71-43-2	2.13	10.43	0.64	0.00	0.09	0.65	0.56
93	1,1,1-Трихлорэтан	71-55-6	2.49	10.23	0.21	0.00	0.00	0.45	0.41
94	Дихлорметан	75-09-2	1.25	6.46	0.30	0.00	-0.54	-0.59	-
95	1,1-Дихлорэтан	75-34-3	1.79	8.30	0.31	0.00	-0.31	-	-
96	трет-Бутанол	75-65-0	0.35	8.75	1.60	-1.28	-1.68	-	-
97	3,3-Диметил-2-бутанон	75-97-8	1.20	11.94	1.88	0.00	0.06	0.06	-
98	Пентахлорэтан	76-01-7	3.22	14.08	0.37	0.00	1.26	1.43	-
99	1-Этинил-циклогексанол	78-27-3	1.73	14.21	1.85	-1.87	-	-0.31	-
100	2-Метил-1-пропанол	78-83-1	0.76	8.75	1.57	-1.15	-1.29	-	-1.25
101	1,2-Дихлорпропан	78-87-5	1.99	10.14	0.76	0.00	-0.01	-0.05	-
102	2-Бутанон	78-93-3	0.29	8.27	1.91	0.00	-1.65	-	-
103	1,1,2-Трихлорэтан	79-00-5	1.89	10.23	0.68	0.00	0.18	0.21	-
104	Трихлорэтен	79-01-6	2.61	10.04	0.22	0.00	0.42	0.47	0.50
105	1,1,2,2-Тетрахлорэтан	79-34-5	2.39	12.16	0.60	0.00	0.77	0.92	-
106	2-Нонанон	821-55-6	3.14	17.44	1.98	0.00	-	0.97	-
107	1,2,3-Трихлорбензол	87-61-6	4.13	16.22	0.24	0.00	1.89	-	2.41

Продолжение табл. 16

№	Соединение	CAS	logK _{ow}	α	ΣC _a	ΣC _d	log(1/LC ₅₀)		
							Pr	Pp	Om
108	Нафталин	91-20-3	3.30	17.70	1.40	0.00	-	1.32	-
109	N,N-Диэтиланилин	91-66-7	3.31	19.25	1.87	0.00	-	0.96	-
110	4-Фенилпиридин	939-23-1	2.59	19.39	3.03	0.00	-	0.98	-
111	2-Ксилол	95-47-6	3.12	14.10	0.63	0.00	1.33	-	1.12
112	1,2-Дихлорбензол	95-50-1	3.43	14.29	0.37	0.00	1.40	1.19	1.97
113	2,4-Дихлортолуол	95-73-8	4.24	16.13	0.56	0.00	1.54	-	-
114	3,4-Дихлортолуол	95-75-0	4.06	16.13	0.39	0.00	1.60	1.74	-
115	1,2,4,5-Тетрахлорбензол	95-94-3	4.60	18.15	0.23	0.00	2.85	-	2.26
116	1,2,3-Трихлорпропан	96-18-4	1.98	12.06	1.17	0.00	0.55	0.41	-
117	3-Пентанон	96-22-0	0.85	10.10	1.88	0.00	-1.26	-	-
118	Ацетофенон	98-86-2	1.58	14.26	2.03	0.00	-0.13	-0.13	-

Ниже представлены межвидовые корреляции значений токсичности:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -0.02(\pm 0.05) + 0.98(\pm 0.02) \log(1/LC_{50})(Pp)_{\text{эксп}} \quad (54)$$

$$n=39; r^2=0.988; s=0.14; F=2982$$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -0.25(\pm 0.08) + 0.95(\pm 0.05) \log(1/LC_{50})(Om)_{\text{эксп}} \quad (55)$$

$$n=27; r^2=0.939; s=0.41; F=386$$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.18(\pm 0.08) + 0.97(\pm 0.05) \log(1/LC_{50})(Om)_{\text{эксп}} \quad (56)$$

$$n=19; r^2=0.957; s=0.31; F=380$$

Полученные уравнения подтверждают хорошо известные корреляции токсичности между тремя видами рыб и показывают возможность построения для них общей QSAR модели. Результаты применения методов LDA и SS для классификации химических соединений на две группы: MOA неполярный наркоз и другие MOA для *P. reticulata* и для всех трех рыб представлены в табл. 17.

Таблица 17

Чувствительность (SN), специфичность (SP) и точность (ACC) при бинарной классификации химических соединений с МОА неполярный наркоз (класс 1) и другие МОА (класс 2) по отношению к *P. reticulata* (Pr), *P. promelas* (Pp) и *O. mykiss* (Om)

№	Объект	Метод	Обучающая выборка				Тестовая выборка			
			n_1/n_2	SN	SP	ACC	n_1/n_2	SN	SP	ACC
1	Pr	LDA_4D	72/163	0.472	0.859	0.740	18/40	0.333	0.850	0.690
2	Pr	LDA_PC	72/163	0.639	0.834	0.774	18/40	0.556	0.825	0.741
3	Pr	SS_1	72/163	0.903	0.914	0.911	18/40	0.889	0.900	0.897
4	Pr	SS_3	72/163	0.986	0.908	0.932	18/40	0.944	0.900	0.914
5	Pr+Pp+Om	LDA_4D	95/244	0.263	0.877	0.705	23/61	0.261	0.934	0.750
6	Pr+Pp+Om	LDA_PC	95/244	0.505	0.857	0.758	23/61	0.348	0.934	0.774
7	Pr+Pp+Om	SS_1	95/244	0.842	0.873	0.864	23/61	0.913	0.869	0.881
8	Pr+Pp+Om	SS_3	95/244	0.811	0.865	0.850	23/61	0.696	0.852	0.810

Примечание: n_1 – число соединений в первом классе, n_2 - число соединений во втором классе.

Линейный дискриминантный анализ, основанный на использовании 4-х HVBOT дескрипторов: α , $\sum q$, $\sum C_a$, $\sum C_d$ (метод LDA_4D в табл. 17) или главных компонент, полученных из 89 дескрипторов (метод LDA_PC), не дал хороших результатов при дискриминации неполярных наркотиков и соединений с другими типами токсического действия.

Применение SS оказывается более эффективным в этом случае. Например, использование SS с учетом только первого ближайшего соседа (SS_1) для обучающей выборки, содержащей 72 соединения с МОА неполярный наркоз и 163 соединения с другими МОА, обеспечивало правильное распознавание 65 соединений в классе 1 (90.3%) и 149 (91.4%) соединений в классе 2 (№ 3, табл. 17). Для тестового ряда соединений

правильная классификация наблюдалась для 16 из 18 неполярных наркотиков (88.9%) и для 36 из 40 (90.0%) соединений с другими МОА.

Подобные результаты были получены для обучающих и тестовых выборок с данными для всех трех рыб. Так, использование SS для обучающей выборки, содержащей 95 соединений с МОА неполярный наркоз и 244 соединения с другими МОА, давало правильное распознавание 80 (84.2%) соединений в первой группе и 213 соединений (87.3%) во второй группе (№ 7, табл. 17). В обучающей выборке правильно классифицированы были 21 из 23 (91.3%) и 53 из 61 (86.9%) соединений для первой и второй группы соответственно.

Обычно токсичность соединений с МОА неполярный наркоз хорошо коррелирует с экспериментальными или рассчитанными величинами $\log K_{ow}$. Однако, как уже было показано нами ранее, коэффициент распределения вещества в системе н-октанол-вода может быть заменен суммой двух других дескрипторов: молекулярной поляризуемостью (α) и свободноэнергетическими протоноакцепторными дескрипторами ($\sum C_a$).

Наш ряд из 90 органических соединений с МОА неполярный наркоз по отношению к *P. reticulata* имеет три очевидных структурных области: хлоралканы и хлорбензолы, которые являются слабыми электронодонорами (соединения 4, 5, 6, 8, 15, 16, 17, 18, 19, 25, 49, 53, 55, 68, 70, 77, 80, 81, 83, 87, 88, 92, 93, 94, 95, 98, 101, 103, 104, 105, 107, 111, 112, 113, 114, 115 и 116 из табл. 16), эфиры и кетоны, которые являются средними электронодонорами (соединения 3, 13, 14, 21, 27, 28, 31, 32, 48, 54, 56, 59, 60, 62, 66, 71, 72, 76, 86, 89, 90, 97, 102, 117 и 118) и соединения с ОН группами (10, 11, 20, 23, 26, 30, 34, 35, 36, 38, 39, 41, 42, 43, 44, 45, 47, 52, 57, 64, 69, 73, 82, 84, 85, 91, 96 и 100). Это дает возможность изучить влияние структурного разнообразия на стабильность QSAR моделей. Соответствующие регрессионные уравнения, основанные на применении $\log K_{ow}$ и HVBOT дескрипторов, представлены ниже.

Слабые электронодоноры:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.86(\pm 0.17) + 0.91(\pm 0.05) \log K_{ow} \quad (57)$$

$$n=37; r^2=0.901; s=0.30; F=317$$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -2.19(\pm 0.06) + 0.257(\pm 0.015) \alpha - 0.48(\pm 0.06) \sum C_a \quad (58)$$

$$n=37; r^2=0.910; s=0.29; F=171$$

Средние электронодоноры:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.78(\pm 0.14) + 0.91(\pm 0.06) \log K_{ow} \quad (59)$$

$$n=25; r^2=0.900; s=0.35; F=207$$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.90(\pm 0.25) + 0.244(\pm 0.015) \alpha - 0.83(\pm 0.12) \sum C_a \quad (60)$$

$$n=25; r^2=0.925; s=0.31; F=136$$

Соединения с ОН группами:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.81(\pm 0.08) + 0.84(\pm 0.04) \log K_{ow} \quad (61)$$

$$n=28; r^2=0.947; s=0.37; F=467$$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -2.46(\pm 0.21) + 0.271(\pm 0.013) \alpha - 0.74(\pm 0.05) \sum C_a \quad (62)$$

$$n=28; r^2=0.961; s=0.32; F=305$$

Так как этот ряд содержит соединения с Н-донорной способностью, соответствующие НУВОТ дескрипторы были включены в корреляцию:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -2.46(\pm 0.25) + 0.271(\pm 0.013) \alpha - 0.74(\pm 0.07) \sum C_a + 0.00(\pm 0.15) \sum C_d \quad (63)$$

$$n=28; r^2=0.961; s=0.33; F=195$$

Очевидно, что протонодонорная способность соединений не вносит вклад в их токсичность.

Уравнения (57), (59) и (61), а также (58), (60) и (62) имеют приблизительно одинаковые коэффициенты. Поэтому три структурные области могут быть объединены, чтобы получить общую модель токсичности для всех соединений с МОА неполярный наркоз по отношению к *P. reticulata*:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.79(\pm 0.06) + 0.88(\pm 0.02) \log K_{ow} \quad (64)$$

$n=90$; $r^2=0.948$; $s=0.34$; $F=1596$; $q^2=0.945$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -2.09(\pm 0.11) + 0.257(\pm 0.008) \alpha - 0.81(\pm 0.03) \sum C_a \quad (65)$$

$n=90$; $r^2=0.956$; $s=0.31$; $F=940$; $q^2=0.952$

Включение в модель дополнительного дескриптора C_d не улучшало статистические характеристики, а соответствующий коэффициент в модели незначимо отличался от нуля.

С целью валидации моделей токсичности, ряд из 90 соединений был разбит на обучающую и тестовую выборки. 72 соединения обучающей выборки имели величины $\log K_{ow}$ в интервале $-1.45 \div 5.32$ с 62 соединениями (86.1%) в интервале $0.00 \div 5.00$. Интервал величин молекулярной поляризуемости составлял $3.25 \div 23.43 \text{ \AA}^3$ (с 91.7% в интервале $5.00 \div 20.00 \text{ \AA}^3$), а интервал суммы Н-акцепторных дескрипторов составлял $0.18 \div 5.77$ (с 93.1% в интервале $0.18 \div 3.12$).

Регрессионные уравнения для обучающего ряда из 72-х соединений с МОА неполярный наркоз по отношению к *P. reticulata* представлены ниже:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.78(\pm 0.07) + 0.88(\pm 0.02) \log K_{ow} \quad (66)$$

$n=72$; $r^2=0.947$; $s=0.35$; $F=1246$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -2.15(\pm 0.12) + 0.262(\pm 0.008) \alpha - 0.80(\pm 0.03) \sum C_a \quad (67)$$

$n=72; r^2=0.960; s=0.31; F=823$

В тестовом ряду находились соединения из всех трех структурных областей (восемь слабых электронодонора, четыре средних электронодонора и шесть гидроксильных соединений) и имели величины $\log K_{ow}$ в интервале - 0.80÷4.04. Интервал величин молекулярной поляризуемости составлял 7.55÷22.08 Å³, а интервал суммы Н-акцепторных дескрипторов составлял 0.28÷3.32.

Использование уравнений (66) и (67) для предсказания расчетных величин токсичности в тестовом ряду дало следующие результаты:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -0.03(\pm 0.07) + 1.01(\pm 0.06) \log(1/LC_{50})(Pr)_{\text{расч}} \quad (68)$$

$n=18; r^2=0.954; s=0.28; F=330$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -0.05(\pm 0.08) + 0.96(\pm 0.06) \log(1/LC_{50})(Pr)_{\text{расч}} \quad (69)$$

$n=18; r^2=0.933; s=0.33; F=222$

Ряд данных токсичности по отношению к *P. promelas* содержал 66 соединений, относящихся к трем рассмотренным выше структурным областям, плюс несколько азотсодержащих производных. Регрессионные модели для обучающей выборки были следующими:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.75(\pm 0.09) + 0.88(\pm 0.03) \log K_{ow} \quad (70)$$

$n=53; r^2=0.927; s=0.33; F=648$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.97(\pm 0.15) + 0.259(\pm 0.011) \alpha - 0.92(\pm 0.03) \sum C_a \quad (71)$$

$n=53; r^2=0.929; s=0.32; F=327$

Использование уравнений (70) и (71) для предсказания расчетных величин токсичности в тестовом ряду дало следующие результаты:

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = 0.09(\pm 0.09) + 1.04(\pm 0.08) \log(1/LC_{50})(Pp)_{\text{расч}} \quad (72)$$

$$n=13; r^2=0.937; s=0.33; F=162$$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = 0.18(\pm 0.07) + 1.03(\pm 0.07) \log(1/LC_{50})(Pp)_{\text{расч}} \quad (73)$$

$$n=13; r^2=0.958; s=0.27; F=251$$

Ряд данных токсичности по отношению к *O. mykiss* содержал только 31 соединение. Регрессионные уравнения для обучающей выборки были следующими:

$$\log(1/LC_{50})(Om)_{\text{эксп}} = -1.56(\pm 0.17) + 0.89(\pm 0.05) \log K_{ow} \quad (74)$$

$$n=25; r^2=0.925; s=0.41; F=284$$

$$\log(1/LC_{50})(Om)_{\text{эксп}} = -1.35(\pm 0.25) + 0.242(\pm 0.017) \alpha - 1.12(\pm 0.09) \sum C_a \quad (75)$$

$$n=25; r^2=0.949; s=0.35; F=203$$

Использование уравнений (74) и (75) для предсказания расчетных величин токсичности в тестовом ряду дало следующие результаты:

$$\log(1/LC_{50})(Om)_{\text{эксп}} = 0.22(\pm 0.20) + 1.13(\pm 0.13) \log(1/LC_{50})(Om)_{\text{расч}} \quad (76)$$

$$n=6; r^2=0.949; s=0.49; F=74.6$$

$$\log(1/LC_{50})(Om)_{\text{эксп}} = 0.23(\pm 0.25) + 0.90(\pm 0.13) \log(1/LC_{50})(Om)_{\text{расч}} \quad (77)$$

$$n=6; r^2=0.923; s=0.61; F=47.7$$

Сходство величин коэффициентов в уравнениях (66), (70) и (74) свидетельствуют о возможности конструирования общей регрессионной модели для *P. reticulata*, *P. promelas* и *O. mykiss* на основе величин $\log K_{ow}$. Обучающий ряд содержал 150 величин токсичности для 89 соединений:

$$\log(1/LC_{50})(Pr+Pp+Om)_{\text{экс}} = -1.75(\pm 0.05) + 0.89(\pm 0.03) \log K_{ow} \quad (78)$$

$n=150; r^2=0.936; s=0.36; F=2171$

Уравнение (78) не является новым с позиции описания базовой токсичности на основе экспериментальных величин $\log K_{ow}$. Однако, общая модель для трех видов рыб позволяет предсказывать величины базовой токсичности для одного из видов на основе экспериментальных величин токсичности для других видов, используя $\log K_{ow}$.

Подобная процедура возможна для создания общей QSAR модели на основе молекулярной поляризуемости и Н-акцепторных дескрипторов:

$$\log(1/LC_{50})(Pr+Pp+Om)_{\text{экс}} = -1.96(\pm 0.09) + 0.258(\pm 0.006) \alpha - 0.88(\pm 0.03) \sum C_a \quad (79)$$

$n=150; r^2=0.943; s=0.34; F=1211$

Статистические характеристики уравнения (79) несколько лучше, чем характеристики уравнения (78). Однако, главное преимущество уравнения (79) состоит в использовании не экспериментальных, а рассчитанных дескрипторов. Кроме того, уравнение (79) дает возможность количественной оценки вкладов с противоположными эффектами: молекулярной поляризуемости соединений, которая увеличивает базовую токсичность, и Н-акцепторной способности, которая уменьшает ее.

Валидация уравнений (78) и (79) была выполнена с использованием комбинации тестовых выборок для *P. reticulata*, *P. promelas* и *O. mykiss*.

Тестовый ряд содержал 37 величин токсичности для 32 соединений, относящихся ко всем структурным областям (хлор- и бромуглеводороды, бензолы, эфиры, кетоны, соединения с ОН группами и азотсодержащие соединения).

Использование уравнений (78) и (79) для предсказания расчетных величин токсичности в тестовом ряду дало следующие результаты:

$$\begin{aligned} \log(1/LC_{50})(Pr+Pp+Om)_{\text{эксп}} &= -0.05(\pm 0.06) \\ &+ 1.02(\pm 0.05) \log(1/LC_{50})(Pr+Pp+Om)_{\text{расч}} \end{aligned} \quad (80)$$

$n=37; r^2=0.930; s=0.36; F=462$

$$\begin{aligned} \log(1/LC_{50})(Pr+Pp+Om)_{\text{эксп}} &= -0.00(\pm 0.05) \\ &+ 0.99(\pm 0.06) \log(1/LC_{50})(Pr+Pp+Om)_{\text{расч}} \end{aligned} \quad (81)$$

$n=37; r^2=0.942; s=0.33; F=564$

Имеется возможность того, что несколько соединений могут присутствовать как в обучающей, так и в тестовой выборке в случае комбинации данных для нескольких видов рыб. Например, число химических соединений в обучающих и тестовых рядах (121), которые были использованы для конструирования уравнений (78) и (79), на три больше, чем общее число соединений, представленных в табл. 16 (118), т.е. три соединения были включены как в обучающий, так и тестовый ряд.

Чтобы избежать появления такой ситуации, ряд из 118 соединений с МОА неполярный наркоз был разбит на обучающую выборку (95 соединений) и тестовую выборку (23 соединения). Все данные по токсичности были включены в регрессионный анализ. Соответствующие уравнения для обучающего ряда представлены ниже:

$$\begin{aligned} \log(1/LC_{50})(Pr+Pp+Om)_{\text{эксп}} &= -1.70(\pm 0.05) + 0.87(\pm 0.02) \log K_{ow} \end{aligned} \quad (82)$$

$n=153; r^2=0.929; s=0.37; F=1976$

$$\log(1/LC_{50})(Pr+Pp+Om)_{\text{эксп}} = -1.98(\pm 0.07) + 0.260(\pm 0.007) \alpha - 0.86(\pm 0.03) \sum C_a \quad (83)$$

$$n=153; r^2=0.943; s=0.33; F=1235$$

Использование уравнений (82) и (83) для предсказания расчетных величин токсичности в тестовом ряду дало следующие результаты:

$$\log(1/LC_{50})(Pr+Pp+Om)_{\text{эксп}} = -0.07(\pm 0.05) + 1.02(\pm 0.03) \log(1/LC_{50})(Pr+Pp+Om)_{\text{расч}} \quad (84)$$

$$n=33; r^2=0.968; s=0.28; F=941$$

$$\log(1/LC_{50})(Pr+Pp+Om)_{\text{эксп}} = -0.11(\pm 0.05) + 1.07(\pm 0.04) \log(1/LC_{50})(Pr+Pp+Om)_{\text{расч}} \quad (85)$$

$$n=33; r^2=0.964; s=0.29; F=842$$

Одно соединение (2,6-диизопропиланилин) было исключено из тестового ряда, так как по литературным данным оно относится к различным МОА.

На рис. 10 представлено распределение химических соединений в обучающем и тестовом ряду с использованием двух дескрипторов (α и $\sum C_a$).

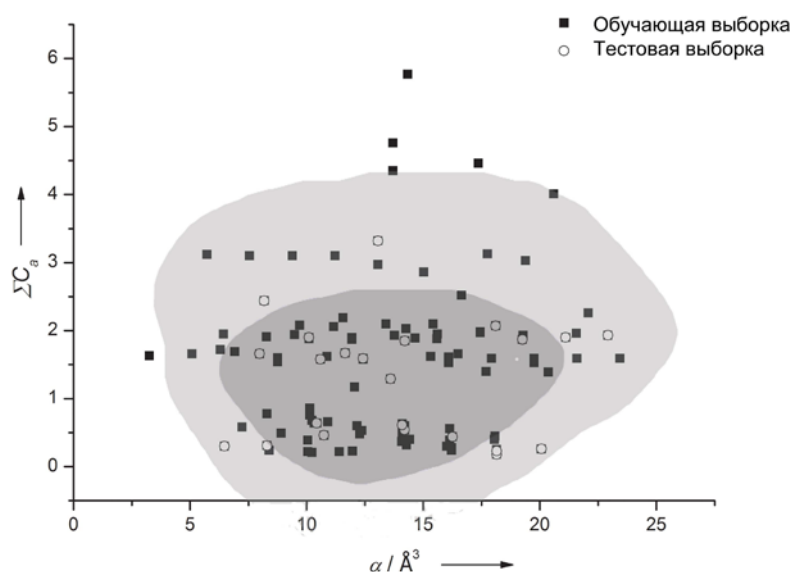


Рис. 10. Распределение химических соединений обучающей и тестовой выборок. Светло-серый контур описывает область, в которой расположены 95% соединений обучающей выборки, темно-серый контур соответствует 75%

Наблюдается ясная кластеризация соединений по отношению к Н-акцепторной способности: слабые акцепторы (ароматика и хлоруглеводороды) находятся в области ΣC_a 0.00÷1.00, средние акцепторы (эфир и кетоны) в области 1.00÷2.50, диолы и алкоксиспирты в области 3.00. Соединения в каждом из этих кластеров имеют широкий диапазон изменения молекулярной поляризуемости (5.00÷25.00 Å³). Следовательно, токсичность соединений в каждом кластере, в первую очередь, является функцией молекулярного объема.

Все соединения тестового ряда находятся в той же области, что и 95% соединений обучающего ряда и их токсичность предсказывается с ошибкой близкой к экспериментальной ошибке.

Хотя модели QSAR, использующие $\log K_{ow}$ как дескриптор, могут быть полезны для предсказательной цели, но они не дают информации о механизме токсического действия просто потому, что $\log K_{ow}$ является композитным дескриптором, включающим в себя несколько физико-химических

дескрипторов, в частности, молекулярный размер и протоноакцепторную способность. Уравнение (83), использующее эти дескрипторы, позволяет высветить механизм действия базовой токсичности. Молекулярный размер увеличивает токсичность, вероятно, из-за увеличения нарушения клеточных мембран, в то время, как Н-акцепторная способность уменьшает токсичность.

4.2. Классификационные и регрессионные модели полярного наркоза

В настоящее время существует две точки зрения относительно механизмов токсичности, связанных с полярным наркозом и базовой токсичностью. Некоторые авторы считают, что имеется реальная разница между этими двумя МОА. Например, в работе [291] выдвинуто предположение о том, что разница между базовой токсичностью и полярным наркозом основана на разнице в физической химии. Базовая токсичность действует благодаря 3-D распределению (способность двигаться во всех направлениях в углеводородоподобных внутренних частях мембран), а полярный наркоз действует благодаря 2-D распределению (включая связывание между функциональной группой наркотика и полярными головными группами фосфатидилхолина на мембранной поверхности). Другие авторы [292] полагают, что нет разницы между рассматриваемыми МОА и что кажущееся различие возникает из-за того, что $\log K_{ow}$ (используемый как параметр базовой токсичности) является неадекватным дескриптором распределения в липидной мембране.

Выше были получены отличные регрессионные модели базовой токсичности органических соединений по отношению к рыбам *P. reticulata* (Pr), *P. promelas* (Prp) и *O. mykiss* (Om) не только в зависимости от коэффициента распределения вещества в системе н-октанол-вода, но и от HVBOT дескрипторов, связанных с молекулярным объемом и протоноакцепторной способностью соединений. Включение в модель дополнительных

дескрипторов, отражающих электростатическое взаимодействие и Н-донорную способность молекул, не улучшало корреляцию.

В табл. 18 представлены расчетные и экспериментальные данные по острой токсичности (LC_{50} , ммоль/л) органических соединений с МОА полярный наркоз по отношению к *P. reticulata*, *P. promelas* и *O. mykiss*, а также величины $\log K_{ow}$ и НУВОТ дескрипторов α ($\text{\AA}^3$), $\sum C_a$, $\sum C_d$.

Таблица 18

Острая токсичность (LC_{50} , ммоль/л), липофильность ($\log K_{ow}$), поляризуемость (α , Å³), сумма свободноэнергетических Н-акцепторных (ΣC_a) и Н-донорных (ΣC_d) дескрипторов органических соединений с МОА полярный наркоз по отношению к *P. reticulata* (Pr), *P. promelas* (Pp) и *O. mykiss* (Om)

№	Соединение	$\log K_{ow}$	α	ΣC_a	ΣC_d	$\log(1/LC_{50})$ (Pr)				$\log(1/LC_{50})$ (Pp)				$\log(1/LC_{50})$ (Om)			
						N	P-N	P	Эксп	N	P-N	P	Эксп	N	P-N	P	Эксп
1	1,2-Диаминопропан	-0.91	8.98	3.98	-5.14	-2.98	1.36	-1.62	-	-3.29	2.21	-1.07	-1.13	-3.28	2.64	-0.64	-
2	1,2-Диметил-3-нитробензол	2.83	15.95	1.36	0.00	0.96	0.20	1.15	1.39	0.96	0.05	1.01	-	1.03	0.84	1.88	-
3	1,2-Диметилпропиламин	1.10	11.30	1.99	-2.57	-0.76	0.79	0.02	-0.51	-0.83	0.99	0.16	-	-0.75	1.45	0.70	-
4	1,3-Диаминопропан	-1.49	8.98	4.18	-5.14	-3.14	1.38	-1.76	-	-3.47	2.28	-1.19	-1.21	-3.47	2.74	-0.74	-
5	1,3-Дигидроксibenзол	0.80	10.98	2.13	-5.11	-0.96	1.07	0.11	0.04	-1.05	1.42	0.38	0.34	-0.96	1.63	0.67	0.04
6	1-Адамантанамина	1.44	18.15	2.15	-1.94	0.89	0.33	1.22	0.78	0.80	0.41	1.21	-	0.81	1.23	2.04	-
7	1-Амино-2-пропанол	0.96	8.27	3.46	-4.08	-2.74	1.25	-1.49	-1.52	-2.99	1.95	-1.04	-	-2.94	2.36	-0.58	-
8	1-Метилгептиламина	2.82	16.81	1.99	-2.57	0.67	0.46	1.13	1.40	0.60	0.55	1.15	-	0.63	1.23	1.86	-
9	1-Нафтол	2.84	17.98	1.98	-2.61	0.98	0.39	1.37	1.50	0.91	0.46	1.37	1.49	0.93	1.18	2.11	-
10	2-(2-Бутоксietокси)этанол	0.56	17.37	4.16	-1.86	-0.94	0.55	-0.40	-0.85	-1.27	1.15	-0.13	-	-1.35	2.26	0.91	-
11	2,2-Диметилпропиламина	1.19	11.30	1.99	-2.57	-0.76	0.79	0.02	-0.74	-0.83	0.99	0.16	-	-0.75	1.45	0.70	-
12	2,3,4,5-Тетрахлоранилин	4.56	19.62	1.05	-3.64	2.16	0.31	2.47	2.81	2.20	0.16	2.36	-	2.26	0.69	2.94	-
13	2,3,4,5-Тетрахлорфенол	3.76	18.42	0.70	-2.40	2.13	0.23	2.36	2.52	2.22	-0.04	2.18	2.75	2.30	0.51	2.81	-
14	2,3,4-Трихлоранилин	3.68	17.70	1.08	-3.64	1.64	0.43	2.07	2.15	1.68	0.32	2.00	1.73	1.75	0.78	2.52	-
15	2,3,5,6-Тетрахлоранилин	4.46	19.62	0.86	-2.57	2.31	0.19	2.50	2.93	2.38	-0.06	2.32	-	2.44	0.55	2.99	-
16	2,3,5,6-Тетрахлорфенол	3.88	18.42	0.70	-2.40	2.13	0.23	2.36	2.26	2.22	-0.04	2.18	-	2.30	0.51	2.81	-
17	2,3,5-Трихлорфенол	3.58	16.49	0.79	-2.40	1.56	0.35	1.91	1.92	1.63	0.15	1.78	-	1.73	0.63	2.36	-
18	2,3,6-Трихлоранилин	3.32	17.70	0.90	-2.57	1.78	0.31	2.09	1.73	1.85	0.11	1.95	-	1.92	0.64	2.57	-
19	2,3,6-Трихлорфенол	3.77	16.49	0.72	-2.32	1.61	0.34	1.95	1.56	1.70	0.11	1.81	-	1.80	0.59	2.39	-
20	2,3,6-Триметилфенол	2.92	16.21	1.50	-1.28	0.91	0.32	1.23	1.21	0.90	0.25	1.15	1.22	0.96	0.95	1.92	-
21	2,3-Дихлорнитробензол	3.05	16.13	1.14	0.00	1.18	0.16	1.35	1.66	1.21	-0.04	1.17	-	1.30	0.72	2.02	-
22	2,4,5-Трихлоранилин	3.69	17.70	1.06	-3.64	1.65	0.43	2.08	1.92	1.70	0.31	2.01	-	1.77	0.77	2.53	-

Продолжение табл. 18

№	Соединение	logK _{ow}	α	ΣC _a	ΣC _d	log(1/LC ₅₀) (Pr)				log(1/LC ₅₀) (Pp)				log(1/LC ₅₀) (Om)			
						N	P-N	P	Эксп	N	P-N	P	Эксп	N	P-N	P	Эксп
23	2,4,5-Трихлорфенол	3.80	16.49	0.67	-2.51	1.65	0.35	2.01	2.20	1.74	0.12	1.86	2.25	1.85	0.58	2.42	2.63
24	2,4,6-Трибромфенол	3.92	18.59	0.72	-1.37	2.16	0.12	2.28	1.70	2.24	-0.19	2.05	-	2.32	0.47	2.79	-
25	2,4,6-Трихлорфенол	4.03	16.49	0.70	-2.21	1.63	0.32	1.96	1.94	1.72	0.09	1.80	1.33	1.82	0.58	2.40	2.54
26	2,4-Дихлоранилин	2.91	15.77	1.19	-3.64	1.05	0.55	1.60	1.41	1.07	0.51	1.59	-	1.16	0.91	2.07	-
27	2,4-Дихлорнитробензол	3.09	16.13	1.10	0.00	1.21	0.16	1.37	1.46	1.25	-0.06	1.19	-	1.33	0.70	2.04	-
28	2,4-Дихлорфенол	3.17	14.56	0.84	-2.48	1.02	0.48	1.50	1.59	1.08	0.33	1.41	1.32	1.20	0.74	1.93	1.80
29	2,4-Диметилфенол	2.30	14.38	1.53	-2.06	0.41	0.51	0.92	0.86	0.40	0.52	0.91	0.87	0.48	1.07	1.55	-
30	2,5-Дихлоранилин	2.92	15.77	1.29	-3.64	0.97	0.56	1.53	1.99	0.98	0.55	1.53	-	1.06	0.96	2.02	-
31	2,5-Дихлорнитробензол	2.90	16.13	1.24	0.00	1.10	0.17	1.27	1.59	1.12	-0.01	1.11	-	1.20	0.77	1.97	-
32	2,5-Дихлорфенол	3.06	14.56	0.94	-2.36	0.93	0.48	1.41	1.58	0.99	0.35	1.34	-	1.10	0.78	1.88	-
33	2,6-Дихлорфенол	2.84	14.56	1.06	-2.02	0.84	0.45	1.29	1.32	0.88	0.34	1.22	-	0.98	0.83	1.81	-
34	2,6-Диметилфенол	2.36	14.38	1.47	-1.58	0.46	0.46	0.92	0.75	0.45	0.43	0.88	-	0.53	1.02	1.56	-
35	2-Аллилфенол	2.55	16.02	1.66	-1.82	0.73	0.40	1.13	0.96	0.70	0.40	1.10	1.01	0.76	1.06	1.82	-
36	2-Амино-5-бромпиридин	1.39	13.83	2.53	-2.52	-0.54	0.68	0.14	-	-0.68	0.97	0.29	-0.01	-0.64	1.61	0.97	-
37	2-Аминоэтанол	-1.31	6.43	3.44	-4.11	-3.20	1.36	-1.84	-1.54	-3.45	2.09	-1.36	-	-3.38	2.43	-0.96	-0.51
38	2-Хлор-4-метиланилин	2.58	15.67	1.72	-3.64	0.59	0.61	1.20	-	0.55	0.70	1.26	0.60	0.61	1.18	1.79	-
39	2-Хлор-4-метилфенол	2.65	14.47	1.06	-2.33	0.81	0.49	1.30	0.60	0.86	0.39	1.25	-	0.96	0.84	1.80	-
40	2-Хлор-4-нитроанилин	2.06	15.68	2.38	-3.64	0.06	0.67	0.73	0.93	-0.06	0.92	0.87	0.93	-0.03	1.51	1.48	-
41	2-Хлор-6-нитротолуол	3.09	16.04	1.36	0.00	0.98	0.19	1.17	1.52	0.99	0.04	1.02	-	1.06	0.84	1.90	-
42	2-Хлоранилин	1.90	13.84	1.60	-3.64	0.21	0.71	0.92	1.31	0.19	0.81	1.00	1.35	0.27	1.19	1.46	-
43	2-Хлорнитробензол	2.24	14.20	1.39	0.00	0.48	0.30	0.78	0.72	0.48	0.20	0.68	-	0.57	0.93	1.49	-
44	2-Хлорфенол	1.94	12.64	1.22	-2.18	0.21	0.60	0.81	1.06	0.23	0.57	0.80	1.05	0.34	0.99	1.34	-
45	2-Цианопиридин	0.50	11.59	3.11	0.00	-1.60	0.61	-0.98	-	-1.80	0.99	-0.81	-0.84	-1.77	1.89	0.12	-
46	2-Этиланилин	1.93	15.58	2.37	-1.00	0.04	0.41	0.45	0.21	-0.07	0.56	0.49	-	-0.05	1.40	1.35	-
47	2-Метоксиэтиламин	-0.67	8.27	3.34	-2.57	-2.65	1.09	-1.55	-0.84	-2.88	1.69	-1.18	-	-2.83	2.24	-0.58	-

Продолжение табл. 18

№	Соединение	logK _{ow}	α	ΣC _a	ΣC _d	log(1/LC ₅₀) (Pr)				log(1/LC ₅₀) (Pp)				log(1/LC ₅₀) (Om)			
						N	P-N	P	Эксп	N	P-N	P	Эксп	N	P-N	P	Эксп
48	2-Метиланилин	1.32	13.75	2.32	-1.00	-0.39	0.51	0.12	0.12	-0.50	0.69	0.19	-	-0.46	1.45	0.99	-
49	2-Метилфенол	1.95	12.54	1.44	-2.27	0.00	0.63	0.64	0.77	0.00	0.66	0.67	0.89	0.10	1.11	1.21	1.11
50	2-Нитроанилин	1.85	13.75	1.81	-2.98	0.02	0.67	0.68	1.15	-0.03	0.79	0.76	-	0.04	1.27	1.32	-
51	2-Нитротолуол	2.30	14.11	1.44	0.00	0.41	0.31	0.73	0.59	0.41	0.22	0.63	-	0.50	0.96	1.45	-
52	2-Фенилфенол	3.09	20.37	2.27	-1.67	1.37	0.18	1.55	1.76	1.27	0.24	1.50	1.44	1.25	1.19	2.43	-
53	2-Трет-бутил-4-метилфенол	3.80	19.88	1.37	-2.19	1.97	0.18	2.15	1.90	1.97	0.04	2.02	-	2.01	0.78	2.78	-
54	3,3-Диметилбутиламин	1.72	13.14	2.06	-2.57	-0.34	0.68	0.34	-0.78	-0.42	0.87	0.45	-	-0.35	1.41	1.05	-
55	3,4,5-Трихлорфенол	4.28	16.49	0.48	-3.52	1.81	0.44	2.24	2.08	1.92	0.20	2.12	-	2.03	0.52	2.55	-
56	3,4-Дихлоранилин	2.69	15.77	1.37	-2.57	0.90	0.46	1.36	1.39	0.91	0.42	1.33	1.33	0.98	0.96	1.94	1.92
57	3,4-Диметилнитробензол	2.91	15.95	1.36	0.00	0.96	0.20	1.15	1.21	0.96	0.05	1.01	-	1.03	0.84	1.88	-
58	3,4-Диметилфенол	2.23	14.38	1.59	-2.28	0.36	0.54	0.90	0.92	0.34	0.57	0.91	-	0.42	1.11	1.53	-
59	3,5-Дихлоранилин	2.90	15.77	1.18	-2.57	1.05	0.45	1.50	1.62	1.08	0.36	1.44	-	1.17	0.86	2.03	-
60	3,5-Дихлорнитробензол	3.13	16.13	1.08	0.00	1.23	0.16	1.39	1.58	1.27	-0.06	1.21	-	1.35	0.69	2.05	-
61	3,5-Дихлорфенол	3.63	14.56	0.58	-3.24	1.23	0.53	1.76	1.78	1.33	0.35	1.67	-	1.45	0.64	2.09	-
62	3-Бензилоксианилин	2.77	23.68	3.12	-2.00	1.54	0.09	1.63	1.34	1.34	0.31	1.64	1.34	1.24	1.49	2.74	-
63	3-Хлоранилин	1.88	13.84	1.68	-2.57	0.15	0.61	0.76	0.98	0.12	0.68	0.80	-	0.19	1.19	1.38	-
64	3-Хлорнитробензол	2.46	14.20	1.32	0.00	0.53	0.30	0.83	1.01	0.54	0.17	0.72	-	0.64	0.89	1.53	-
65	3-Хлорфенол	2.59	12.64	0.93	-2.91	0.44	0.65	1.09	1.30	0.50	0.57	1.07	-	0.63	0.88	1.50	-
66	3-Циано-4,6-диметил-2-гидроксипиридин	2.03	15.53	3.67	-2.27	-1.02	0.66	-0.37	-	-1.30	1.18	-0.11	-0.03	-1.33	2.10	0.77	-
67	3-Этиланилин	1.88	15.58	1.92	-2.13	0.41	0.48	0.89	0.65	0.35	0.56	0.91	-	0.39	1.22	1.62	-
68	3-Метоксифенол	1.58	12.82	1.98	-2.57	-0.36	0.70	0.34	0.22	-0.43	0.87	0.44	0.22	-0.36	1.38	1.02	-
69	3-Метиланилин	1.44	13.75	2.26	-2.00	-0.35	0.61	0.26	0.47	-0.45	0.81	0.36	-	-0.40	1.46	1.06	-
70	3-Метилфенол	1.96	12.54	1.35	-2.40	0.08	0.64	0.72	0.48	0.08	0.65	0.74	-	0.19	1.07	1.26	1.08
71	3-Нитроанилин	1.37	13.75	2.40	-3.52	-0.46	0.77	0.31	0.24	-0.58	1.07	0.49	-	-0.53	1.59	1.06	-
72	3-Нитрофенол	2.00	12.55	1.84	-3.45	-0.32	0.79	0.47	1.07	-0.37	0.96	0.60	-	-0.29	1.36	1.07	-

Продолжение табл. 18

№	Соединение	logK _{ow}	α	ΣC _a	ΣC _d	log(1/LC ₅₀) (Pr)				log(1/LC ₅₀) (Pp)				log(1/LC ₅₀) (Om)			
						N	P-N	P	Эксп	N	P-N	P	Эксп	N	P-N	P	Эксп
73	3-Нитротолуол	2.42	14.11	1.39	0.00	0.45	0.31	0.76	0.65	0.46	0.20	0.66	-	0.55	0.93	1.48	-
74	3-Трифторметил-4-нитрофенол	3.00	14.11	2.08	-4.03	-0.11	0.77	0.67	-	-0.19	1.00	0.82	1.36	-0.13	1.44	1.31	-
75	4-(N-метоксиметил)аминофенол	0.48	16.49	3.30	-4.42	-0.48	0.78	0.30	-0.73	-0.70	1.28	0.58	-	-0.73	1.97	1.24	-
76	4-Ацетилпиридин	0.48	13.55	3.34	0.00	-1.27	0.52	-0.75	-	-1.50	0.91	-0.59	-0.14	-1.51	1.93	0.42	-
77	4-Амино-2-нитрофенол	0.96	14.03	2.88	-6.52	-0.78	1.10	0.32	0.64	-0.95	1.63	0.68	0.63	-0.93	1.94	1.00	-
78	4-Броманилин	2.26	14.54	1.70	-2.40	0.31	0.55	0.86	0.56	0.28	0.61	0.89	0.56	0.35	1.16	1.51	-
79	4-Бутиланилин	3.05	19.25	2.17	-1.94	1.16	0.26	1.42	1.16	1.07	0.33	1.40	1.17	1.07	1.19	2.26	-
80	4-Хлор-2-нитротолуол	3.05	16.04	1.40	0.00	0.95	0.19	1.14	1.44	0.95	0.05	1.00	-	1.02	0.86	1.88	-
81	4-Хлор-3,5-диметилфенол	3.45	16.31	1.36	-2.97	1.05	0.47	1.52	1.66	1.06	0.43	1.49	-	1.12	0.95	2.07	2.31
82	4-Хлор-3-метилфенол	3.10	14.47	1.05	-2.71	0.82	0.53	1.35	1.33	0.87	0.44	1.30	1.42	0.97	0.85	1.82	-
83	4-Хлоранилин	1.88	13.84	1.65	-2.56	0.17	0.60	0.78	0.67	0.14	0.67	0.82	0.61	0.22	1.17	1.40	1.06
84	4-Хлорнитробензол	2.39	14.20	1.42	0.00	0.45	0.31	0.76	1.42	0.45	0.21	0.66	-	0.54	0.94	1.48	-
85	4-Хлорфенол	2.48	12.64	1.00	-2.88	0.39	0.65	1.04	1.18	0.44	0.59	1.03	1.32	0.56	0.91	1.47	1.83
86	4-Дециланилин	6.09	30.26	2.27	-1.95	3.94	-0.39	3.55	3.58	3.84	-0.52	3.32	3.58	3.72	0.80	4.52	-
87	4-Этоксид-2-нитроанилин	2.38	17.70	2.78	-3.28	0.26	0.55	0.81	0.85	0.10	0.85	0.95	0.85	0.08	1.61	1.69	-
88	4-Этиланилин	1.92	15.58	2.24	-2.42	0.15	0.54	0.69	0.52	0.05	0.71	0.76	0.22	0.08	1.39	1.47	-
89	4-Этилфенол	2.58	14.38	1.32	-2.48	0.58	0.53	1.11	1.07	0.59	0.51	1.10	-	0.68	0.98	1.67	-
90	4-Гексилоксианилин	3.64	23.20	2.65	-1.70	1.80	0.05	1.84	1.78	1.65	0.14	1.79	-	1.58	1.27	2.85	-
91	4-Метоксифенол	1.34	12.82	2.07	-2.38	-0.43	0.69	0.25	0.05	-0.51	0.87	0.36	0.05	-0.44	1.42	0.97	0.64
92	4-Метиланилин	1.39	13.75	2.32	-1.96	-0.39	0.61	0.22	0.72	-0.50	0.82	0.32	-0.17	-0.46	1.49	1.03	-
93	4-Метилфенол	1.94	12.54	1.44	-2.45	0.00	0.65	0.66	0.74	0.00	0.69	0.69	-	0.10	1.12	1.22	1.14
94	4-н-Бутилфенол	3.56	18.05	1.43	-2.47	1.44	0.32	1.77	1.47	1.44	0.25	1.69	-	1.49	0.89	2.38	-
95	4-Нитроанилин	1.39	13.75	2.39	-3.54	-0.45	0.77	0.32	0.23	-0.57	1.07	0.50	-	-0.52	1.59	1.06	-
96	4-Нитробензамид	0.82	15.74	3.57	-4.16	-0.89	0.82	-0.07	-	-1.15	1.40	0.25	0.10	-1.18	2.12	0.94	-
97	4-Нитрофенол	1.96	12.55	1.81	-3.57	-0.29	0.80	0.50	0.99	-0.34	0.97	0.63	0.49	-0.26	1.35	1.09	1.25

Продолжение табл. 18

№	Соединение	logK _{ow}	α	ΣC _a	ΣC _d	log(1/LC ₅₀) (Pr)				log(1/LC ₅₀) (Pp)				log(1/LC ₅₀) (Om)			
						N	P-N	P	Эксп	N	P-N	P	Эксп	N	P-N	P	Эксп
98	4-Нитротолуол	2.37	14.11	1.39	0.00	0.45	0.31	0.76	0.67	0.46	0.20	0.66	-	0.55	0.93	1.48	-
99	4-Нонилфенол	6.21	27.22	1.43	-2.47	3.83	-0.23	3.60	3.20	3.83	-0.49	3.34	-	3.78	0.53	4.31	-
100	4-н-Пентилфенол	4.09	19.88	1.43	-2.47	1.92	0.21	2.13	2.12	1.92	0.10	2.02	-	1.95	0.82	2.77	-
101	4-Октиланилин	5.03	26.59	2.27	-1.95	2.98	-0.17	2.82	3.23	2.88	-0.22	2.66	-	2.80	0.95	3.75	-
102	4-Феноксифенол	3.35	20.64	2.01	-2.45	1.65	0.22	1.87	1.58	1.58	0.24	1.81	1.58	1.57	1.08	2.65	-
103	4-Фенилазофенол	3.96	22.43	3.29	-3.02	1.08	0.28	1.36	2.24	0.85	0.61	1.46	-	0.76	1.67	2.43	-
104	4-Пиколин	1.22	11.56	2.21	0.00	-0.87	0.54	-0.34	-	-0.97	0.69	-0.28	-0.64	-0.90	1.44	0.55	-
105	4-Пропилфенол	3.20	16.21	1.45	-2.50	0.95	0.44	1.39	1.09	0.95	0.41	1.35	-	1.01	0.98	1.99	-
106	4-трет-Бутилфенол	3.31	18.05	1.34	-2.41	1.52	0.31	1.83	1.46	1.53	0.21	1.74	1.46	1.58	0.84	2.42	-
107	4-трет-Пентилфенол	3.83	19.88	1.38	-2.46	1.96	0.21	2.17	1.81	1.97	0.08	2.05	1.80	2.00	0.79	2.79	-
108	6-Хлор-2-пиридинол	1.78	11.93	2.98	-2.27	-1.40	0.81	-0.59	-	-1.59	1.24	-0.35	-0.22	-1.56	1.90	0.35	-
109	α,α,α,4-тетрафтор-2-метиланилин	2.51	13.38	1.55	-1.00	0.13	0.47	0.60	0.78	0.12	0.46	0.57	-	0.21	1.08	1.29	-
110	α,α,α,4-тетрафтор-3-метиланилин	2.51	13.38	1.87	-2.13	-0.13	0.61	0.48	0.77	-0.18	0.72	0.54	-	-0.11	1.29	1.18	-
111	Амиламин	1.49	11.30	1.96	-2.57	-0.74	0.79	0.05	-0.31	-0.80	0.98	0.18	-	-0.72	1.43	0.72	-
112	Анилин	0.94	11.91	2.27	-2.08	-0.83	0.73	-0.10	-0.09	-0.93	0.97	0.04	-0.05	-0.87	1.54	0.67	-
113	Бензиламин	1.09	13.62	2.43	-2.57	-0.52	0.69	0.17	0.02	-0.64	0.96	0.32	-	-0.60	1.57	0.98	-
114	Бутиламин	0.97	9.47	2.00	-2.57	-1.27	0.93	-0.34	-0.56	-1.32	1.14	-0.18	-	-1.21	1.52	0.31	-
115	Катехол	0.88	10.98	2.08	-3.00	-0.92	0.86	-0.06	-	-1.00	1.11	0.11	1.08	-0.91	1.52	0.61	1.09
116	Дециламин	4.10	20.48	1.99	-2.57	1.62	0.24	1.86	2.18	1.55	0.26	1.81	-	1.55	1.08	2.63	-
117	Диэтиламин	0.58	9.47	1.94	-1.50	-1.20	0.79	-0.41	-	-1.26	0.97	-0.29	-	-1.15	1.45	0.30	-0.40
118	Додециламин	5.16	24.15	1.99	-2.57	2.58	0.02	2.59	3.27	2.51	-0.04	2.47	-	2.47	0.93	3.40	-
119	Этилацетат	0.73	8.54	1.69	0.00	-1.24	0.67	-0.57	-	-1.27	0.75	-0.52	-	-1.14	1.30	0.16	-0.74
120	Этиламин	-0.13	5.80	1.87	-2.57	-2.13	1.12	-1.01	-0.70	-2.15	1.39	-0.76	-	-2.00	1.61	-0.40	-
121	Гептиламин	2.57	14.97	1.99	-2.57	0.18	0.62	0.80	0.72	0.12	0.70	0.82	-	0.17	1.30	1.47	-
122	Гексиламин	2.06	13.14	1.96	-2.57	-0.28	0.72	0.45	0.25	-0.33	0.84	0.51	-	-0.26	1.36	1.10	-

Продолжение табл. 18

№	Соединение	logK _{ow}	α	ΣC_a	ΣC_d	log(1/LC ₅₀) (Pr)				log(1/LC ₅₀) (Pp)				log(1/LC ₅₀) (Om)			
						N	P-N	P	Эксп	N	P-N	P	Эксп	N	P-N	P	Эксп
123	Нитробензол	1.85	12.28	1.32	0.00	0.03	0.41	0.45	-0.03	0.05	0.33	0.37	0.01	0.16	0.97	1.13	-
124	Нониламин	3.57	18.64	1.99	-2.57	1.14	0.42	1.56	1.82	1.08	0.41	1.48	-	1.09	1.15	2.24	-
125	Октиламин	3.04	16.81	1.99	-2.57	0.66	0.52	1.18	1.40	0.60	0.55	1.15	-	0.63	1.23	1.86	-
126	Пентафторанилин	1.86	11.46	1.61	-2.57	-0.41	0.74	0.33	0.69	-0.44	0.85	0.41	-	-0.33	1.25	0.92	-
127	Фенол	1.46	10.71	1.27	-2.49	-0.33	0.75	0.42	0.45	-0.32	0.78	0.47	0.46	-0.19	1.11	0.92	1.19
128	Пропиламин	0.48	7.63	2.00	-2.57	-1.75	1.03	-0.72	-0.72	-1.80	1.29	-0.51	-	-1.67	1.60	-0.07	-
129	Пиридин	0.65	9.73	2.21	0.00	-1.35	0.65	-0.71	-	-1.45	0.83	-0.61	-0.10	-1.35	1.52	0.16	1.24
130	Хинолин	2.03	17.00	2.47	0.00	0.33	0.23	0.56	0.63	0.20	0.34	0.54	0.45	0.21	1.36	1.56	-
131	втор-Бутиламин	0.74	9.47	1.98	-2.57	-1.25	0.92	-0.33	-0.58	-1.30	1.14	-0.16	-	-1.19	1.51	0.32	-
132	Трет-Октиламин	2.69	16.81	2.04	-2.57	0.62	0.53	1.15	0.72	0.55	0.57	1.12	-	0.58	1.25	1.83	-
133	Тридециламин	5.10	25.98	1.99	-2.57	3.06	0.02	3.08	3.46	2.98	-0.18	2.80	-	2.92	0.86	3.78	-
134	Ундециламин	4.63	22.31	1.99	-2.57	2.10	0.22	2.32	2.91	2.03	0.11	2.14	-	2.01	1.01	3.01	-

Примечание: N – неполярный наркоз, P – полярный наркоз (расчетные значения токсичности).

Нами были проанализированы межвидовые корреляции токсичности для трех видов рыб. Соответствующие регрессионные уравнения представлены ниже:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = 0.13(\pm 0.06) + 0.94(\pm 0.05) \log(1/LC_{50})(Pp)_{\text{эксп}} \quad (86)$$

$n=37; r^2=0.912; s=0.23; F=362$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -0.52(\pm 0.12) + 1.02(\pm 0.08) \log(1/LC_{50})(Om)_{\text{эксп}} \quad (87)$$

$n=15; r^2=0.929; s=0.26; F=170$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.04(\pm 0.21) + 0.89(\pm 0.13) \log(1/LC_{50})(Om)_{\text{эксп}} \quad (88)$$

$n=12; r^2=0.745; s=0.32; F=29.2$

Уравнения (86)-(88) подобны соответствующим уравнениям (54)-(56) для соединений с базовой токсичностью. При этом может быть сделано аналогичное заключение о возможности конструирования единой модели токсичности для всех трех рыб. Также нужно отметить, что качество токсикологических данных, особенно для *P. promelas* и *O. mykiss*, не является совершенным. Это связано, в первую очередь, с тем, что они были получены в разных лабораториях.

В настоящей работе мы использовали три типа бинарной классификации: дискриминацию между соединениями с МОА полярный наркоз (P) и другими МОА (M), дискриминацию между соединениями с МОА полярный наркоз (P) и неполярный наркоз (N), дискриминацию между соединениями с неспецифической токсичностью (полярный и неполярный наркоз) (P+N) и другими МОА (M). Результаты применения методов LDA и SS для вышеуказанных классификаций представлены в табл. 19.

Таблица 19

Чувствительность (SN), специфичность (SP) и точность (ACC) при бинарной классификации органических соединений с различными MOA (класс 1/класс 2) по отношению к *P. reticulata* (Pr) и *P. promelas* (Pp)

№	Рыба	MOA	Метод	Обучающая выборка				Тестовая выборка			
				n_1/n_2	SN	SP	ACC	n_1/n_2	SN	SP	ACC
1	Pr	P/M	LDA_4D	96/138	0.792	0.899	0.855	23/34	0.826	0.882	0.860
2	Pr	P/M	LDA_PC	96/138	0.833	0.899	0.872	23/34	0.870	0.882	0.877
3	Pr	P/M	SS_1	96/138	0.896	0.913	0.906	23/34	0.957	0.941	0.947
4	Pr+Pp	P/M	LDA_4D	106/280	0.623	0.925	0.842	26/70	0.654	0.871	0.812
5	Pr+Pp	P/M	LDA_PC	106/280	0.594	0.921	0.832	26/70	0.538	0.857	0.771
6	Pr+Pp	P/M	SS_1	106/280	0.868	0.904	0.894	26/70	0.923	0.914	0.917
7	Pr	P/N	LDA_4D	96/72	0.823	0.833	0.827	23/18	0.870	0.833	0.854
8	Pr	P/N	LDA_PC	96/72	0.958	0.944	0.952	23/18	0.957	0.944	0.951
9	Pr	P/N	SS_1	96/72	0.958	0.944	0.952	23/18	1.000	1.000	1.000
10	Pr+Pp	P/N	LDA_4D	106/148	0.764	0.892	0.839	26/36	0.846	0.833	0.839
11	Pr+Pp	P/N	LDA_PC	106/148	0.849	0.905	0.882	26/36	0.885	0.944	0.919
12	Pr+Pp	P/N	SS_1	106/148	0.962	0.919	0.937	26/36	0.962	0.861	0.903
13	Pr	(P+N)/M	LDA_4D	168/66	0.964	0.318	0.782	41/16	0.976	0.188	0.754
14	Pr	(P+N)/M	LDA_PC	168/66	0.958	0.455	0.816	41/16	0.927	0.250	0.737
15	Pr	(P+N)/M	SS_1	168/66	0.952	0.667	0.872	41/16	0.951	0.563	0.842
16	Pr+Pp	(P+N)/M	LDA_4D	262/124	0.947	0.274	0.731	65/31	0.985	0.194	0.729
17	Pr+Pp	(P+N)/M	LDA_PC	262/124	0.947	0.387	0.767	65/31	0.908	0.323	0.719
18	Pr+Pp	(P+N)/M	SS_1	262/124	0.897	0.718	0.839	65/31	1.000	0.710	0.906

Примечание: n_1 – число соединений в первом классе, n_2 - число соединений во втором классе; MOA: P - полярный наркоз, N – неполярный наркоз, M – другие.

Линейный дискриминантный анализ, основанный на применении 4-х НУВОТ дескрипторов: α , q^+ , $\sum C_a$, $\sum C_d$, (метод LDA_4D в табл. 19), привел к получению удовлетворительной дискриминации между полярными наркотиками и соединениями с другими MOA. 76 из 96 полярных наркотиков в

обучающей выборке для *P. reticulata* были распознаны правильно с использованием этого метода (№ 1, табл. 19). Для обучающей выборки *P. reticulata* + *P. promelas* LDA_4D (№ 4, табл. 19) показывает более плохой результат (66 правильно распознанных соединений из 106).

Результаты классификации с помощью LDA на 10-12 главных компонентах, полученных с использованием 89 дескрипторов (метод LDA_PC в табл. 19), улучшались незначительно: в случае *P. reticulata* доля правильного распознавания соединений с MOA полярный наркоз составляла 83.3% и 87.0% в обучающем и тестовом ряду соответственно. Для комбинированных данных (*P. reticulata* + *P. promelas*) были получены несколько худшие результаты как в обучающей, так и тестовой выборке. В случае соединений с другими MOA этот метод обеспечивал правильное распознавание на уровне 89.9% в обучающей выборке для *P. reticulata* и 92.1% для комбинированных данных. Для тестовой выборки эти проценты составили 88.2 и 85.7% соответственно.

Использование для дискриминации соединений метода SS на основе первого структурного соседа (SS_1) приводило к получению более лучших результатов. Например, применение SS_1 для обучающего ряда, содержащего 96 соединений с MOA полярный наркоз и 138 соединений с другими MOA по отношению к *P. reticulata*, давало правильное распознавание 86 (89.6%) полярных наркотиков и 126 (91.3%) соединений с другими MOA (№ 3, табл. 19). В тестовом ряду была проведена успешная классификация 22 из 23 (95.7%) соединений в первом классе и 32 из 34 (94.1%) во втором классе.

Подобные результаты были получены для обучающих и тестовых выборок, содержащих данные для *P. reticulata* и *P. promelas*. Так, использование SS_1 для обучающей выборки, содержащей 106 соединений с MOA полярный наркоз и 280 соединений с другими MOA (№ 6, табл. 19), обеспечило правильную классификацию 92 (86.8%) соединений в первом классе и 253 (90.4%) соединений во втором классе. В тестовой выборке корректная

дискриминация выполнена для 24 из 26 (92.3%) соединений с МОА полярный наркоз и для 64 из 70 (91.4%) соединений с другими МОА.

Дискриминация между соединениями с МОА полярный и неполярный наркоз важна для дискуссии, касающейся существования разницы между этими двумя типами токсического действия. Здесь метод LDA_4D снова не дал хороших результатов: только 79 из 96 (82.3%) соединений с МОА полярный наркоз были правильно классифицированы в обучающей выборке *P. reticulata* (№ 7, табл. 19). В случае комбинированных данных для *P. reticulata* и *P. promelas* результаты были хуже (81 из 106 (76.4%)). Корректная классификация соединений с МОА неполярный наркоз в обучающей выборке составляла 83.3% для *P. reticulata* и 89.2% для *P. reticulata* и *P. promelas*.

Метод LDA_PC показал более хорошие результаты: 95.8% правильно распознанных полярных наркотиков в обучающей выборке для *P. reticulata* и 84.5% для *P. reticulata* и *P. promelas*. Корректная классификация соединений с МОА неполярный наркоз в обучающей выборке для *P. reticulata* составляла 94.4% и 90.5% для комбинированных данных *P. reticulata* и *P. promelas*. В тестовой выборке для *P. reticulata* доля правильного распознавания составляла 95.7% в первом классе и 94.4% во втором классе. В случае комбинированных данных было получено 88.5% и 94.4% для первого и второго класса соответственно.

Похожие хорошие результаты были получены с помощью SS_1 метода. В случае обучающего ряда соединений для *P. reticulata* 95.8% полярных наркотиков и 94.4% неполярных наркотиков были правильно проклассифицированы. Более того, в тестовом ряду для *P. reticulata* все соединения с МОА полярный и неполярный наркоз были распознаны правильно. Для совместных данных *P. reticulata* и *P. promelas* результаты были несколько хуже, но все еще отличными: в обучающем ряду степень правильной классификации составляла 96.2% и 91.9%, в тестовом ряду 96.2% и 86.1% для соединений с МОА полярный и неполярный наркоз соответственно.

Все три использованных метода показали хорошие результаты при классификации соединений с неспецифической токсичностью (полярный и неполярный наркоз). В случае *P. reticulata* степень правильного распознавания при использовании методов LDA_4D, LDA_PC, SS_1 составила в обучающей выборке 96.4%, 95.8%, 95.2%, а в тестовой выборке 97.6%, 92.7% и 95.1%. В случае комбинированных данных *P. reticulata* и *P. promelas* в обучающей выборке были получены величины распознавания с помощью трех методов 94.7%, 94.7%, 93.3%, а в тестовой выборке 98.5%, 90.8% и 100.0%.

Однако, ситуация с классификацией соединений со специфической токсичностью была значительно хуже. Методы LDA_4D и LDA_PC не могли обеспечить правильное распознавание большинства соединений. Метод SS_1 показал долю корректной классификации этих соединений в обучающем ряду для *P. reticulata* на уровне 66.7% и 71.8% для совместных данных *P. reticulata* и *P. promelas*. В тестовом ряду эти величины составили, соответственно, 56.3% и 71.0%.

В целом, использованные методы (особенно SS_1) обеспечили правильную классификацию химических соединений с МОА полярный и неполярный наркоз.

Использованный нами ряд из 119 органических соединений с МОА полярный наркоз по отношению к *P. reticulata* имеет несколько очевидных структурных областей: соединения без Н-донорной способности (15 нитробензолов, № в табл. 18: 2, 21, 27, 31, 41, 43, 51, 57, 60, 64, 73, 80, 84, 98, 123), соединения с Н-донорной способностью (24 амина: 1, 3, 4, 6, 8, 11, 47, 54, 111, 113, 114, 116-118, 120-122, 124, 125, 128, 131-134), (36 анилинов: 12, 14, 15, 18, 22, 26, 30, 38, 40, 42, 46, 48, 50, 56, 59, 62, 63, 67, 69, 71, 75, 77-79, 83, 86-88, 90, 92, 95, 101, 109, 110, 112, 126), (46 фенолов: 5, 9, 13, 16, 17, 19, 20, 23-25, 28, 29, 32-35, 39, 44, 49, 52, 53, 55, 58, 61, 65, 68, 70, 72, 75, 77, 81, 82, 85, 89, 91, 93, 94, 97, 99, 100, 102, 103, 105-107, 127). Это дает возможность исследовать влияние структурного разнообразия на стабильность моделей QSAR.

Соответствующие регрессионные уравнения, основанные на $\log K_{ow}$ и HYBOT дескрипторах, представлены ниже.

Нитробензолы:

$$\log(1/LC_{50})(Pr)_{эксп} = -1.86(\pm 0.40) + 1.12(\pm 0.15) \log K_{ow} \quad (89)$$

$$n=15; r^2=0.813; s=0.22; F=56.5$$

$$\log(1/LC_{50})(Pr)_{эксп} = -4.13(\pm 1.38) + 0.365(\pm 0.056) \alpha - 0.17(\pm 0.58) \sum C_a \quad (90)$$

$$n=15; r^2=0.830; s=0.22; F=29.3$$

Амины:

$$\log(1/LC_{50})(Pr)_{эксп} = -1.11(\pm 0.16) + 0.80(\pm 0.06) \log K_{ow} \quad (91)$$

$$n=24; r^2=0.883; s=0.51; F=166$$

$$\log(1/LC_{50})(Pr)_{эксп} = -2.36(\pm 0.60) + 0.239(\pm 0.017) \alpha - 0.27(\pm 0.20) \sum C_a \quad (92)$$

$$n=24; r^2=0.921; s=0.43; F=123$$

$$\log(1/LC_{50})(Pr)_{эксп} = -2.66(\pm 0.65) + 0.243(\pm 0.018) \alpha - 0.34(\pm 0.21) \sum C_a - 0.15(\pm 0.14) \sum C_d \quad (93)$$

$$n=24; r^2=0.926; s=0.43; F=83.3$$

Анилины:

$$\log(1/LC_{50})(Pr)_{эксп} = -0.70(\pm 0.11) + 0.74(\pm 0.04) \log K_{ow} \quad (94)$$

$$n=36; r^2=0.906; s=0.30; F=327$$

$$\log(1/LC_{50})(Pr)_{эксп} = -0.33(\pm 0.27) + 0.192(\pm 0.014) \alpha - 0.87(\pm 0.09) \sum C_a \quad (95)$$

$$n=36; r^2=0.892; s=0.33; F=136$$

$$\begin{aligned} \log(1/LC_{50})(Pr)_{\text{эксп}} = & -0.67(\pm 0.31) + 0.196(\pm 0.013) \alpha - 0.88(\pm 0.08) \sum C_a \\ & - 0.10(\pm 0.05) \sum C_d \end{aligned} \quad (96)$$

$n=36; r^2=0.905; s=0.31; F=101$

Фенолы:

$$\begin{aligned} \log(1/LC_{50})(Pr)_{\text{эксп}} = & -0.52(\pm 0.11) + 0.63(\pm 0.04) \log K_{ow} \end{aligned} \quad (97)$$

$n=46; r^2=0.876; s=0.26; F=312$

$$\begin{aligned} \log(1/LC_{50})(Pr)_{\text{эксп}} = & -0.29(\pm 0.25) + 0.162(\pm 0.014) \alpha - 0.70(\pm 0.08) \sum C_a \end{aligned} \quad (98)$$

$n=46; r^2=0.839; s=0.30; F=112$

$$\begin{aligned} \log(1/LC_{50})(Pr)_{\text{эксп}} = & -0.68(\pm 0.27) + 0.171(\pm 0.013) \alpha - 0.80(\pm 0.08) \sum C_a \\ & - 0.15(\pm 0.05) \sum C_d \end{aligned} \quad (99)$$

$n=46; r^2=0.863; s=0.28; F=88.3$

Близость коэффициентов в соответствующих регрессионных уравнениях для всех четырех структурных областей дает возможность построить модель токсичности для всех полярных наркотиков.

Все полярные наркотики:

$$\begin{aligned} \log(1/LC_{50})(Pr)_{\text{эксп}} = & -0.87(\pm 0.07) + 0.76(\pm 0.03) \log K_{ow} \end{aligned} \quad (100)$$

$n=119; r^2=0.880; s=0.35; F=860$

$$\begin{aligned} \log(1/LC_{50})(Pr)_{\text{эксп}} = & -0.85(\pm 0.16) + 0.199(\pm 0.008) \alpha - 0.70(\pm 0.05) \sum C_a \end{aligned} \quad (101)$$

$n=119; r^2=0.882; s=0.35; F=435$

$$\begin{aligned} \log(1/LC_{50})(Pr)_{\text{эксп}} = & -1.04(\pm 0.16) + 0.202(\pm 0.008) \alpha - 0.74(\pm 0.05) \sum C_a \\ & - 0.10(\pm 0.03) \sum C_d \end{aligned} \quad (102)$$

$n=119; r^2=0.894; s=0.33; F=323$

Добавление в качестве дескриптора $\sum q^-$ к уравнениям (89)-(102) не улучшало их статистические характеристики.

С целью валидации моделей токсичности 119 соединений с МОА полярный наркоз были разбиты на обучающую и тестовую выборки. 96 соединений обучающей выборки имели величины $\log(1/LC_{50})$ в интервале -1.54÷3.58, при этом 80 соединений (83.3%) в интервале 0.04÷2.93; величины $\log K_{ow}$ лежали в интервале -1.31÷6.21 с 84 соединениями (87.5%) в интервале 0.48÷3.92. Интервал величин молекулярной полярности составлял 6.40÷30.26 Å³ (с 91.6% соединений в интервале 10.71÷23.68 Å³), интервал суммы протоноакцепторных дескрипторов составлял 0.48÷4.16 (с 89.6% соединений в интервале 1.08÷3.12) и интервал суммы протонодонорных дескрипторов составлял -6.52÷0.00 (с 85.4% соединений в интервале -4.42÷-1.00). Результаты QSAR моделирования представлены ниже:

Обучающая выборка:

$$\log(1/LC_{50})(Pr)_{\text{экс}} = -0.88(\pm 0.08) + 0.76(\pm 0.03) \log K_{ow} \quad (103)$$

$$n=96; r^2=0.887; s=0.35; F=736$$

$$\log(1/LC_{50})(Pr)_{\text{экс}} = -0.81(\pm 0.17) + 0.201(\pm 0.009) \alpha - 0.74(\pm 0.05) \sum C_a \quad (104)$$

$$n=96; r^2=0.894; s=0.34; F=394$$

$$\log(1/LC_{50})(Pr)_{\text{экс}} = -1.02(\pm 0.16) + 0.205(\pm 0.008) \alpha - 0.79(\pm 0.05) \sum C_a - 0.11(\pm 0.03) \sum C_d \quad (105)$$

$$n=96; r^2=0.910; s=0.31; F=312$$

В состав 23-х соединений тестовой выборки входили соединения из всех структурных областей и имели величины $\log(1/LC_{50})$ в интервале -1.84÷3.23, величины $\log K_{ow}$ лежали в интервале -0.67÷5.03. Интервал величин

молекулярной поляризуемости составлял $8.27 \div 26.59 \text{ \AA}^3$, интервал суммы протоноакцепторных дескрипторов составлял $0.70 \div 3.34$ и интервал суммы протонодонорных дескрипторов составлял $-3.64 \div 0.00$.

Использование уравнений (103)-(105) для предсказания величин $\log(1/LC_{50})$ соединений тестового ряда дало следующие результаты:

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -0.00(\pm 0.14) + 0.99(\pm 0.09) \log(1/LC_{50})(Pr)_{\text{расч}} \quad (106)$$

$$n=23; r^2=0.842; s=0.38; F=112$$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -0.02(\pm 0.14) + 0.96(\pm 0.09) \log(1/LC_{50})(Pr)_{\text{расч}} \quad (107)$$

$$n=23; r^2=0.840; s=0.38; F=111$$

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = 0.02(\pm 0.14) + 0.92(\pm 0.09) \log(1/LC_{50})(Pr)_{\text{расч}} \quad (108)$$

$$n=23; r^2=0.834; s=0.39; F=106$$

Ряд данных по острой токсичности органических соединений с МОА полярный наркоз по отношению к *P. promelas* содержал 50 соединений, относящихся к структурным областям, рассмотренным выше. Регрессионные уравнения для этих соединений были следующими:

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.60(\pm 0.11) + 0.64(\pm 0.05) \log K_{ow} \quad (109)$$

$$n=50; r^2=0.806; s=0.41; F=200$$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.52(\pm 0.26) + 0.164(\pm 0.014) \alpha - 0.58(\pm 0.06) \sum C_a \quad (110)$$

$$n=50; r^2=0.843; s=0.37; F=126$$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.87(\pm 0.25) + 0.169(\pm 0.012) \alpha - 0.61(\pm 0.06) \sum C_a - 0.13(\pm 0.03) \sum C_d \quad (111)$$

$$n=50; r^2=0.883; s=0.33; F=116$$

Добавление в состав независимых переменных $\sum q^-$ не улучшало статистических характеристик уравнений.

Ряд из 50 соединений был разделен на обучающую и тестовую выборки. Для обучающего ряда были получены следующие регрессионные уравнения:

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.60(\pm 0.12) + 0.64(\pm 0.05) \log K_{ow} \quad (112)$$

$n=40; r^2=0.799; s=0.44; F=161$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.57(\pm 0.31) + 0.167(\pm 0.017) \alpha - 0.57(\pm 0.07) \sum C_a \quad (113)$$

$n=40; r^2=0.840; s=0.39; F=97.2$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.88(\pm 0.28) + 0.172(\pm 0.014) \alpha - 0.62(\pm 0.06) \sum C_a - 0.14(\pm 0.04) \sum C_d \quad (114)$$

$n=40; r^2=0.884; s=0.34; F=91.4$

Использование уравнений (112)-(114) для предсказания величин $\log(1/LC_{50})$ соединений тестового ряда дало следующие результаты:

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.05(\pm 0.23) + 1.03(\pm 0.18) \log(1/LC_{50})(Pp)_{\text{расч}} \quad (115)$$

$n=10; r^2=0.804; s=0.32; F=32.8$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.18(\pm 0.23) + 1.07(\pm 0.18) \log(1/LC_{50})(Pp)_{\text{расч}} \quad (116)$$

$n=10; r^2=0.824; s=0.30; F=37.5$

$$\log(1/LC_{50})(Pp)_{\text{эксп}} = -0.22(\pm 0.19) + 1.05(\pm 0.14) \log(1/LC_{50})(Pp)_{\text{расч}} \quad (117)$$

$n=10; r^2=0.876; s=0.25; F=56.7$

Ряд данных по токсичности по отношению к *O. mykiss* содержал только 19 соединений. Регрессионные модели, полученные для этого ряда, были следующими:

$$\log(1/LC_{50})(Om)_{\text{эксп}} = -0.08(\pm 0.19) + 0.66(\pm 0.08) \log K_{ow} \quad (118)$$

$n=19; r^2=0.786; s=0.47; F=62.3$

$$\log(1/LC_{50})(Om)_{\text{эксп}} = -1.88(\pm 1.21) + 0.269(\pm 0.066) \alpha - 0.21(\pm 0.28) \sum C_a \quad (119)$$

$n=19; r^2=0.782; s=0.48; F=28.7$

$$\log(1/LC_{50})(Om)_{\text{эксп}} = -1.89(\pm 1.30) + 0.270(\pm 0.075) \alpha - 0.21(\pm 0.31) \sum C_a - 0.00(\pm 0.11) \sum C_d \quad (120)$$

$n=19; r^2=0.782; s=0.50; F=17.9$

Близость величин коэффициентов уравнений (100) и (109) сигнализирует о возможности создания общей регрессионной модели для *P. reticulata* и *P. promelas*, основанной на величинах $\log K_{ow}$. Соответствующий ряд содержал 169 значений токсичности:

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = -0.78(\pm 0.06) + 0.72(\pm 0.02) \log K_{ow} \quad (121)$$

$n=169; r^2=0.860; s=0.37; F=1022$

Подобная процедура возможна и для конструирования общей модели QSAR для *P. reticulata* и *P. promelas* на основе молекулярной поляризуемости и протонноакцепторных и протонодонорных дескрипторов:

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = -1.02(\pm 0.14) + 0.192(\pm 0.007) \alpha - 0.68(\pm 0.04) \sum C_a - 0.11(\pm 0.02) \sum C_d \quad (122)$$

$n=169; r^2=0.886; s=0.34; F=427$

Статистические характеристики уравнения (122) немного лучше, чем у уравнения (121). Однако, главным преимуществом модели (122) является использование расчетных, а не экспериментальных дескрипторов. Кроме того, эта модель позволяет количественно оценить вклад противоположных эффектов молекулярной поляризуемости и Н-донорной способности, которые увеличивают токсичность соединений, и Н-акцепторной способности, которая уменьшает ее.

Комбинированный ряд данных, содержащий 169 величин $\log(1/LC_{50})$ для 132 соединений, был разделен на обучающую и тестовую выборки. Обучающий ряд состоял из 138 величин токсичности для 106 соединений, включая 64 соединения с токсичностью по отношению только к *P. reticulata*, 10 – к *P. promelas* и 32 – к *P. reticulata* и *P. promelas* (соединения № в табл. 18: 1, 3, 4, 6-8, 10-21, 23-32, 34-38, 40, 41, 43-46, 48-52, 54-56, 58-64, 67-71, 73-78, 79, 80, 82-85, 87-95, 97, 100, 102-113, 116, 118, 120-123, 126, 129-131, 133, 134). Для обучающего ряда были получены следующие модели токсичности:

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = -0.88(\pm 0.07) + 0.75(\pm 0.03) \log K_{ow} \quad (123)$$

$$n=138; r^2=0.867; s=0.35; F=887$$

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = -0.89(\pm 0.17) + 0.197(\pm 0.010) \alpha - 0.66(\pm 0.04) \sum C_a \quad (124)$$

$$n=138; r^2=0.855; s=0.37; F=398$$

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = -1.09(\pm 0.17) + 0.198(\pm 0.009) \alpha - 0.70(\pm 0.04) \sum C_a - 0.11(\pm 0.02) \sum C_d \quad (125)$$

$$n=138; r^2=0.874; s=0.34; F=309$$

Валидация уравнений (123)-(125) была выполнена с использованием тестового ряда, который состоял из 31 величины токсичности для 26 соединений, включая

18 соединений с токсичностью по отношению только к *P. reticulata*, 3 – к *P. promelas* и 5 – к *P. reticulata* и *P. promelas* (соединения № в табл. 18: 2, 5, 9, 22, 33, 39, 42, 47, 53, 57, 65, 66, 72, 81, 86, 96, 98, 99, 101, 114, 115, 124, 125, 127, 128, 132):

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = 0.26(\pm 0.10) + 0.86(\pm 0.06) \log(1/LC_{50})(Pr+Pp)_{\text{расч}} \quad (126)$$

$n=31$; $r^2=0.881$; $s=0.40$; $F=215$

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = 0.26(\pm 0.08) + 0.89(\pm 0.05) \log(1/LC_{50})(Pr+Pp)_{\text{расч}} \quad (127)$$

$n=31$; $r^2=0.911$; $s=0.35$; $F=297$

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = 0.21(\pm 0.08) + 0.91(\pm 0.05) \log(1/LC_{50})(Pr+Pp)_{\text{расч}} \quad (128)$$

$n=31$; $r^2=0.918$; $s=0.33$; $F=326$

Рис. 11 демонстрирует распределение химических соединений из обучающей и тестовой выборки в пространстве из трех дескрипторов ($\sum C_a$, $\sum C_d$).

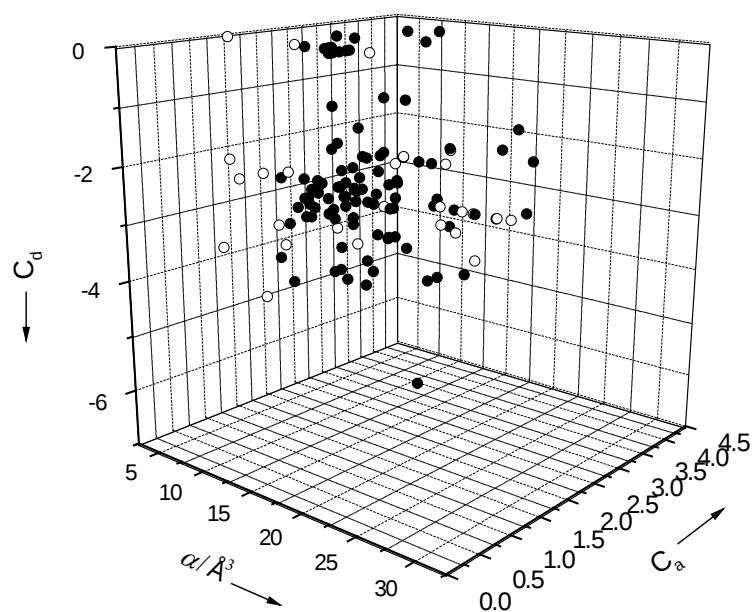


Рис. 11. Расположение органических соединений обучающей (●) и тестовой (○) выборок с МОА полярный наркоз по отношению к *P. reticulata* и *P. promelas* в пространстве из трех дескрипторов: молекулярная поляризуемость (α), свободноэнергетические Н-акцепторные дескрипторы ($\sum C_a$) и Н-донорные дескрипторы ($\sum C_d$)

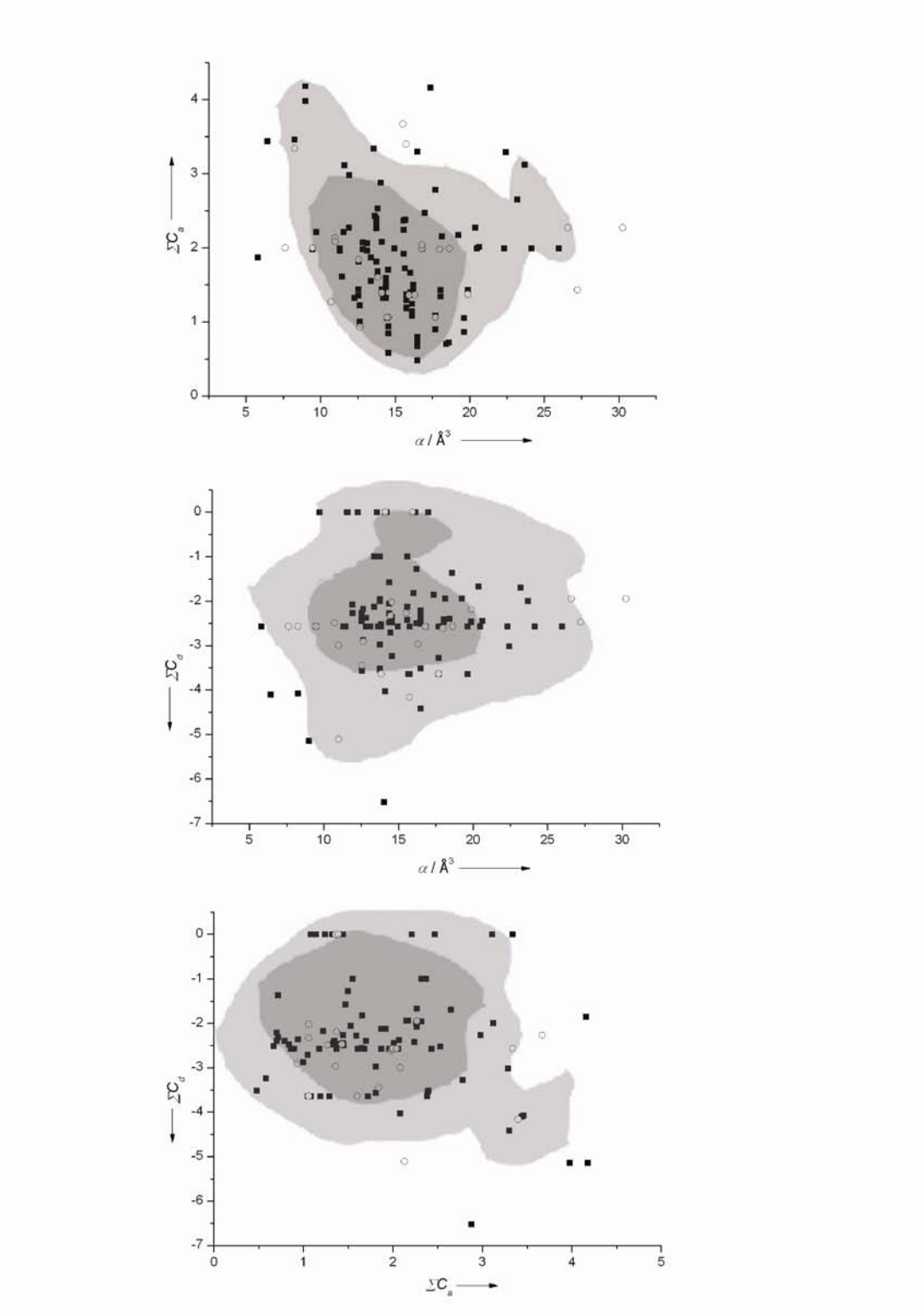


Рис. 12. Распределение химических соединений обучающей (◆) и тестовой (○) выборок. Светло-серый контур описывает область, в которой расположены 95% соединений обучающей выборки, темно-серый контур соответствует 75%

На рис. 12 представлены проекции химических соединений на три пары дескрипторов. Наблюдается отчетливая кластеризация химических соединений по отношению к Н-донорной способности. Анилины расположены в области $\sum C_d$ от -2.80 до -4.00, фенолы в интервале -1.80÷-2.60, нитробензолы имеют $\sum C_d=0.00$. Соединения в каждом из этих кластеров имеют широкий диапазон изменения молекулярной поляризуемости от 5.00 до 25.00 Å³. Следовательно, изменения токсичности соединений внутри каждого кластера связаны, главным образом, с величиной α . Почти все соединения из тестового ряда расположены в 95% области обучающего ряда и их токсичности предсказаны с ошибками близкими к экспериментальным ошибкам.

Модели QSAR, использующие $\log K_{ow}$ как дескриптор, могут быть полезны для предсказательных целей. Например, как это видно из уравнения (121), токсичность соединений с МОА полярный наркоз коррелирует хорошо (хотя не так хорошо, как в случае неполярного наркоза) с $\log K_{ow}$. Более того, сравнение, например, уравнения (121) с соответствующим уравнением для неполярного наркоза ($\log(1/LC_{50}) = -1.70(\pm 0.05) + 0.88(\pm 0.02) \log K_{ow}$) показывает, что соединения с одинаковой величиной $\log K_{ow}$, в случае МОА полярный наркоз, имеют более высокие значения токсичности, чем в случае МОА неполярный наркоз, до величин $\log K_{ow}$ достигающих 7. Тем не менее, дескриптор $\log K_{ow}$ не дает информации о механизме действия, потому что он является композитным дескриптором, включающим в себя несколько физико-химических дескрипторов. С другой стороны, модели, использующие HVBOT дескрипторы для описания неполярного и полярного наркоза, позволяют заглянуть внутрь механизма неспецифической токсичности. Целесообразно сравнить соответствующие уравнения.

P. reticulata:

полярный наркоз (уравнение (102)):

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -1.04(\pm 0.16) + 0.202(\pm 0.008) \alpha - 0.74(\pm 0.05) \sum C_a \\ - 0.10(\pm 0.03) \sum C_d$$

неполярный наркоз (уравнение (65)):

$$\log(1/LC_{50})(Pr)_{\text{эксп}} = -2.09(\pm 0.11) + 0.257(\pm 0.008) \alpha - 0.81(\pm 0.12) \sum C_a$$

«чистый» эффект полярного наркоза (полярный наркоз – неполярный наркоз):

$$\Delta \log(1/LC_{50})(Pr)_{\text{эксп}} = 1.05(\pm 0.27) - 0.055(\pm 0.016) \alpha + 0.07(\pm 0.17) \sum C_a \\ - 0.10(\pm 0.03) \sum C_d \quad (129)$$

P. reticulata + *P. promelas*:

полярный наркоз (уравнение (122)):

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = -1.02(\pm 0.14) + 0.192(\pm 0.007) \alpha - 0.68(\pm 0.04) \sum C_a \\ - 0.11(\pm 0.02) \sum C_d$$

неполярный наркоз (уравнение (130)):

$$\log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = -2.01(\pm 0.09) + 0.256(\pm 0.006) \alpha - 0.84(\pm 0.02) \sum C_a \quad (130)$$

«чистый» эффект полярного наркоза (полярный наркоз – неполярный наркоз):

$$\Delta \log(1/LC_{50})(Pr+Pp)_{\text{эксп}} = 0.99(\pm 0.23) - 0.064(\pm 0.013) \alpha + 0.16(\pm 0.06) \sum C_a \\ - 0.11(\pm 0.02) \sum C_d \quad (131)$$

Интересно отметить, что уравнения (129) и (131), связанные с эффектом чистого полярного наркоза, очень близки к уравнению растворимости в н-октанолу [381]:

$$\log S_0 = 1.06(\pm 0.32) - 0.063(\pm 0.013) \alpha + 0.03(\pm 0.07) \sum C_a \\ - 0.14(\pm 0.08) \sum C_d \quad (132)$$

$$n=23; r^2=0.957; s=0.16; F=69.0$$

Из представленных уравнений (129) и (131) ясно, что рост поляризуемости молекул ведет к уменьшению разницы токсичностей, вычисленных по модели полярного и неполярного наркоза, в то время как рост числа Н-акцепторных и

Н-донорных центров в молекуле приводит к увеличению этой разницы. Полученный результат вместе с заметным вкладом протонодонорного дескриптора $\sum C_d$ в полярный наркоз, свидетельствует о существовании различия между соединениями, относящимися по типу токсического действия к неполярному и полярному наркозу и обусловленному дополнительным межмолекулярным взаимодействием с клеточными мембранами [291]. Вклады неполярного наркоза (рассчитанные с использованием уравнений (65), (71), (75) для *P. reticulata*, *P. promelas* и *O. mykiss*, соответственно) и эффектов полярного наркоза (рассчитанные с использованием уравнений (129), (разность (111) и (71)), (131) для *P. reticulata*, *P. promelas* и *O. mykiss*, соответственно) в общую токсичность всех изученных соединений представлены в табл. 18. Вклад «чистого» эффекта полярного наркоза в общую токсичность совершенно заметен (особенно для соединений, содержащих amino и нитрогруппы) и зависит от гидрофобности соединений. Например, в случае аминопроводных возрастание длины углеводородной цепи сопровождается уменьшением вклада «чистого» эффекта полярного наркоза в токсичность соединений по отношению к *P. reticulata* (1.12 для этиламина, 1.03 для пропиламина, 0.93 для бутиламина, 0.72 для гексиламина, 0.62 для гептиламина, 0.53 для октиламина, 0.42 для нониламина и 0.22 для ундециламина). Такое поведение может быть объяснено возрастанием тенденции соединения двигаться внутрь углеводородоподобной части мембраны с возрастанием его гидрофобности. Это также объясняет почему соединения с высокой гидрофобностью обладают только базовой токсичностью.

4.3. Линейный дискриминантно-регрессионный подход

Ранее уже указывалось, что эффективным способом для построения QSAR моделей острой токсичности является предварительная классификация соединений в соответствии с их MOA или с их принадлежностью к тому или

иному химическому классу. Для реализации такого подхода широко используются такие методы, как LDA, MLR, PLS, SVM, ANN.

Следует отметить, что указанные подходы классификации молекул основаны на априори известной информации о принадлежности молекул к тому или иному МОА или химическому классу. При отсутствии информации о МОА и присутствии неопределенности по отнесению к тому или иному химическому классу (например, в случае полифункциональных соединений) для группировки химических соединений используются другие методы исследования, в частности, кластерный анализ. При этом результат определяется, главным образом, выбранной метрикой для расчета матрицы расстояний, способом включения или исключения молекул и методом определения числа значимых кластеров.

Кроме того, следует подчеркнуть, что традиционный подход для изучения связи структура – токсичность предусматривает сначала классификацию молекул, а затем построение количественных моделей для выделенных классов. В принципе, возможен другой порядок действий: сначала выбор модели количественной связи токсичности с молекулярными характеристиками в общем виде (например, в виде регрессионной зависимости с установленным числом и видом переменных), а затем нахождение с помощью определенных процедур соответствующих этой модели молекулярных кластеров.

Как уже отмечалось ранее, при изучении острой токсичности химических соединений, широкое признание получила концепция неспецифической токсичности или наркоза. Концепция наркоза включает в себя существование простых (чаще всего, линейных) зависимостей между токсичностью химических соединений и их коэффициентами распределения в системе н-октанол - вода. Учитывая, что около половины промышленных токсикантов относится к неспецифическому типу действия, задача дискриминации таких соединений от всех остальных представляется актуальной.

В настоящей работе в качестве тестового массива данных, включая структуры, LC_{50} и $\log K_{ow}$, был использован хорошо охарактеризованный ряд из 293-х органических соединений, обладающих острой токсичностью по отношению к *P. reticulata*, в который входят соединения с 4 типами МОА: неполярный наркоз, полярный наркоз, реактивный, особый [16]. Целью работы являлось изучение возможности применения комбинации дискриминантного и регрессионного анализа для выделения кластеров химических соединений, обладающих неспецифической токсичностью по отношению к *P. reticulata*, и создания регрессионных моделей острой токсичности для них, без использования априорной информации о типе токсического действия каждого соединения.

Для реализации линейного дискриминантно – регрессионного подхода (ЛДРП) была создана компьютерная программа, алгоритм которой можно описать в виде последовательности ряда шагов:

- 1) определение вида регрессионной зависимости (в нашем случае это модель $\log(1/LC_{50}) = a + b \cdot \log K_{ow}$, где a , b – коэффициенты);
 - 2) размещение всех N химических соединений в соответствующем пространстве ($\log(1/LC_{50})$ и $\log K_{ow}$);
 - 3) выбор стартовой комбинации молекул (две молекулы i и j);
 - 4) построение регрессионной модели для полученного кластера;
 - 5) определение ближайшей (по нормали) молекулы к линии регрессии;
 - 6) присоединение найденной молекулы к предыдущему кластеру;
 - 7) повторение шагов 4 - 6 до появления в кластере $(N-1)$ молекул включительно;
 - 8) повторение шагов 3 - 7 для другой стартовой комбинации молекул.
- Таким образом, для N исследуемых молекул потребуется $N \cdot (N-1)/2$ итераций, при этом на каждой итерации формируется $(N-2)$ кластеров, а общее число полученных кластеров составит $N \cdot (N-1) \cdot (N-2)/2$.

Следующим шагом при использовании ЛДРП является анализ полученных кластеров с целью отбора кластеров, отвечающих целевым характеристикам. Для этого был разработан ряд критериев, которые определялись для каждого кластера в отдельности. Первый критерий (МСС*) характеризует качество бинарной классификационной модели, отделяющей рассматриваемый кластер молекул от всех остальных молекул, и основанный на величине МСС. В качестве классификатора в настоящей работе использовался LDA. Вторым критерий (М*) связан с размером определенного паттерна, выявляемого при анализе формы зависимости МСС от размера кластера N. Во внимание принимались только такие участки кривой (и соответствующие им кластеры), которые имели максимальные значения МСС. При этом число точек слева и справа от максимума должно быть не менее, чем $(M^*-1)/2$, где M^* - общий размер паттерна (нечетное число). Еще одним критерием служила величина коэффициента корреляции (R^*), которая характеризовала степень линейности кластеров молекул.

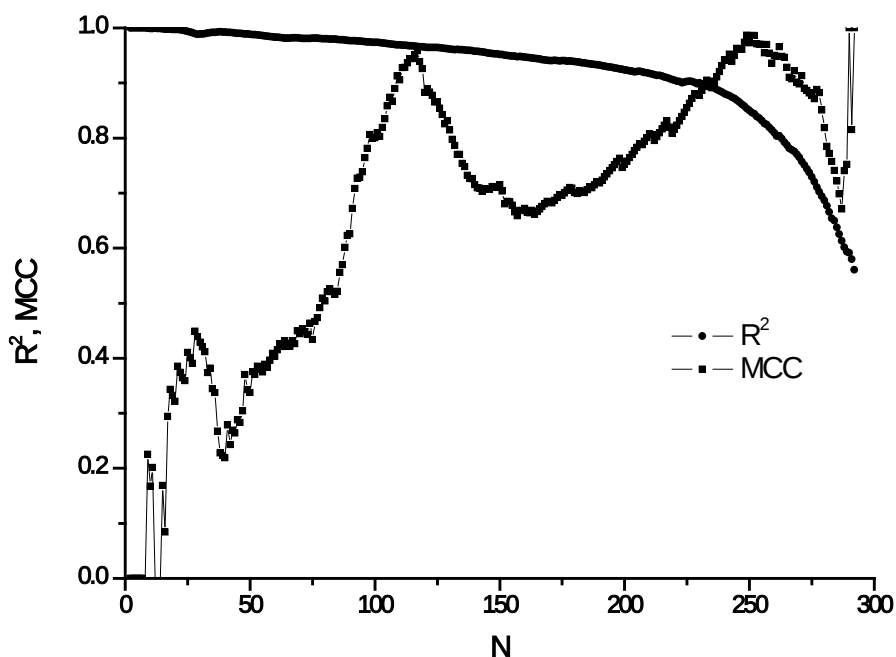


Рис. 13. Зависимость линейного коэффициента корреляции и корреляционного коэффициента Мэтьюза от числа молекул в кластере

На рис. 13 представлена типичная картина, наблюдаемая при использовании ЛДРП. При увеличении числа молекул в кластере коэффициент линейной корреляции регрессионных уравнений плавно уменьшается от 1.000 (2 молекулы) до 0.560 (292 молекулы). MCC ведет себя более сложным образом. Минимальная величина для стартовой пары молекул, линейно неотделимых от остальных молекул, составляет 0.000, а при размере кластера 292 молекулы $MCC=1.000$. При этом при увеличении размера кластера наблюдается несколько кластерных пиков, размер которых и положение зависит от стартовой молекулярной пары.

Для дальнейшего исследования полученных $293 \cdot 292 \cdot 291 / 2 = 12448398$ молекулярных кластеров был проведен их отбор с использованием следующих граничных значений критериев: $M^* \geq 25$, $R^{2*} \geq 0.9$, $MCC^* \geq 0.9$.

Таблица 20

Статистические характеристики выделенных кластеров при использовании ЛДРП

№	i	j	N	R ²	MCC
1	14	35	117	0.966	0.958
2	21	51	116	0.967	0.931
3	22	183	111	0.970	0.920
4	23	269	105	0.970	0.920
5	31	208	110	0.970	0.940
6	40	133	123	0.946	0.904
7	41	76	115	0.967	0.944
8	42	76	117	0.961	0.917
9	42	83	116	0.966	0.951
10	42	87	116	0.966	0.951
11	71	87	117	0.959	0.944
12	94	133	120	0.940	0.902

Примечание. i, j – номера стартовой пары молекул.

Из данных табл. 20 следует, что в результате отбора было выявлено двенадцать кластеров. Минимальный размер кластера равен 105 молекулам, а максимальный – 123 молекулам. Относительно небольшой диапазон изменения размеров позволил выдвинуть предположение о близости составов полученных групп. Действительно, при анализе содержимого двенадцати кластеров была обнаружена большая группа молекул (N=80), которая входит в состав всех кластеров.

С целью отнесения молекул выявленной общей группы к тому или иному типу действия были привлечены литературные данные [16, 238, 310]. В работе [16] приведены сведения об изначальном распределении 293-х молекул по МОА. Так, 90 соединений по типу действия отнесены к неполярному наркозу, 121 – к полярному наркозу, 51 соединение отмечены как реактивные и 31 – как особые. В результате для выделенной группы из 80-ти молекул было установлено, что 53 молекулы относятся к группе с МОА неполярный наркоз,

19 – к полярному наркозу, 5 молекул являются реактивными и 3 – особыми. Т.е. основную массу молекул составляют молекулы с неспецифическим характером действия (неполярный наркоз + полярный наркоз). Их доля оказывается равной $100 \cdot (53+19)/80=90\%$. Оставшиеся 8 молекул лишь только с некоторой вероятностью могут быть отнесены в группу со специфическим (реактивный + особый) характером действия. Три из них являются хлорзамещенными фенолами, а одна молекула представляет собой простой эфир. Как следует из вышеприведенных литературных источников, эти химические соединения могут относиться к группам с различным МОА. Так, 2,3,4,6 – тетрахлорфенол по одним данным обладает МОА неполярный наркоз [238], а по другим данным относится к реактивным соединениям [16]. В работе [310] пентахлорфенол входит в состав группы с МОА полярный наркоз, в то время как в работе [16] он относится к группе с реактивным МОА. Таким образом, можно считать, что выделенная группа соединений по размеру близка к группе с изначально установленным МОА неполярный наркоз и практически полностью состоит из химических соединений с неспецифическим характером действия, большинство из которых составляют молекулы с МОА неполярный наркоз.

Учитывая, что анализируемый в настоящей работе метод ЛДРП не требует наличия априори заданных классов соединений для обучения, для сопоставимого сравнения была проведена кластеризация 293 исследованных химических соединений с помощью НСА. Принимая во внимание, что ЛДРП использовался с применением двух переменных: $\log K_{ow}$ и $\log(1/LC_{50})$, кластеризация также проводилась в этом пространстве переменных. Для проверки достоверности результатов НСА были использованы литературные данные о принадлежности исследуемых молекул к тому или иному МОА [16]. Результаты расчетов представлены в табл. 21.

Таблица 21

Число и состав выделенных кластеров при использовании НСА

Метод	m	Кластер 1				Кластер 2				Кластер 3				Кластер 4			
		N _I	N _{II}	N _{III}	N _{IV}	N _I	N _{II}	N _{III}	N _{IV}	N _I	N _{II}	N _{III}	N _{IV}	N _I	N _{II}	N _{III}	N _{IV}
1	2	75	118	48	28	15	3	3	3	-	-	-	-	-	-	-	-
2	2	54	59	32	7	36	62	19	24	-	-	-	-	-	-	-	-
3	4	90	121	51	27	0	0	0	2	0	0	0	1	0	0	0	1
4	2	42	82	26	25	48	39	25	6	-	-	-	-	-	-	-	-
5	4	52	102	44	17	8	8	4	11	30	11	3	3	0	0	0	1
6	2	42	84	33	26	48	37	18	5	-	-	-	-	-	-	-	-
7	2	68	89	35	8	22	32	16	23	-	-	-	-	-	-	-	-

Примечание. Методы образования кластеров: 1 – межгрупповые связи, 2 – внутригрупповые связи, 3 – ближайший сосед, 4 – дальний сосед, 5 – центроидная кластеризация, 6 – медианная кластеризация, 7 – метод Варда; m – число выделенных кластеров; N_I, N_{II}, N_{III}, N_{IV} – число молекул с МОА неполярный наркоз, полярный наркоз, реактивный, особый соответственно.

Анализ полученных результатов показывает, что число выделенных кластеров колеблется в пределах от 2-х до 4-х. Минимальный размер кластера равен 1, а максимальный составляет 289 молекул. При этом состав кластеров оказывается очень неоднородным. Почти все кластеры (за исключением кластеров 2, 3 и 4, метод ближайшего соседа и кластера 4, центроидная кластеризация), содержат молекулы, относящиеся ко всем четырем МОА. Ни один из выделенных кластеров по своему размеру и содержанию не приближается к заданному изначально распределению молекул по 4 типам действия, в том числе, к группе из 90 молекул с МОА неполярный наркоз.

Как уже отмечалось выше, использование ЛДРП позволяет выделить группу молекул, которая по размеру и составу близка к группе соединений с МОА неполярный наркоз. Учитывая это, мы сконструировали для этого кластера регрессионную модель зависимости токсичности от липофильности:

$$\log(1/LC_{50}) = -1.81(\pm 0.06) + 0.88(\pm 0.02) \log K_{ow} \quad (133)$$

$$n=80; r^2=0.966; s=0.23; F=2199; q^2=0.964$$

Статистические характеристики уравнения (133) свидетельствуют о его хорошей описательной и прогностической способности.

Ранее нами для исследуемого ряда из 293-х химических соединений на основе известного распределения молекул по МОА была получена регрессионная зависимость острой токсичности 90 соединений с типом действия неполярный наркоз от $\log K_{ow}$ (уравнение (64)):

$$\log(1/LC_{50}) = -1.79(\pm 0.06) + 0.88(\pm 0.02) \log K_{ow}$$

$$n=90; r^2=0.948; s=0.34; F=1596; q^2=0.945$$

Как уже выше указывалось, для химических соединений с МОА неполярный наркоз характерно существование линейных зависимостей между токсичностью и $\log K_{ow}$. Выделенная нами группа, в основном, состоит из таких соединений. Поэтому не вызывает удивления практически полное совпадение параметров уравнений (133) и (64).

Нами было показано, что $\log K_{ow}$ можно представить в виде линейной комбинации молекулярной полярности суммы свободноэнергетических дескрипторов Н-связи $\sum C_a$. Преимуществом такого подхода является использование дескрипторов, рассчитываемых непосредственно из структуры химического соединения. С учетом вышесказанного уравнение (133) можно преобразовать в следующую модель:

$$\log(1/LC_{50}) = -1.97(\pm 0.11) + 0.245(\pm 0.007) \alpha - 0.79(\pm 0.04) \sum C_a \quad (134)$$

$$n=80; r^2=0.945; s=0.30; F=665.2; q^2=0.940$$

Аналогично, из уравнения (64) получается модель (65):

$$\log(1/LC_{50}) = -2.09(\pm 0.11) + 0.257(\pm 0.008) \alpha - 0.81(\pm 0.03) \sum C_a$$

$$n=90; r^2=0.956; s=0.31; F=940; q^2=0.952$$

Как и следовало ожидать, параметры уравнений (134) и (65) оказываются близкими друг к другу.

4.4. Ретардантная активность солей четвертичного аммония

Четвертичные аммониевые соли сыграли выдающуюся роль в качестве регуляторов роста и развития растений и стресспротекторов при создании интенсивных биотехнологий основных продовольственных и технических сельскохозяйственных культур [422]. Однако развитие синтетической, физической и биологической химии этих соединений имеет более общее системное значение. Выявление количественных зависимостей физиологической активности низкомолекулярных биорегуляторов от их структуры и физико-химических свойств является одним из средств компьютерного молекулярного дизайна и прогноза физиологической активности новых соединений.

В табл. 22 приведены экспериментальные данные по антигиббереллиновой активности и расчетные значения физико-химических параметров для ряда производных холина и триэтилхолина. На их основе были получены следующие регрессионные модели:

$$I = 38.3(\pm 10.2) + 5.64(\pm 0.74) \alpha - 68.7(\pm 5.5) C_a^{\max} \quad (135)$$

$$n=6; r^2=0.982; s=6.0; q^2=0.724$$

$$I = 4.04(\pm 0.33) \alpha - 19.0(\pm 6.2) C_a^{\max} + 13.9(\pm 4.2) C_d \quad (136)$$

$$n=6; r^2=0.961; s=6.4; q^2=0.706$$

Из представленных данных ясно, что решающую роль в формировании антигиббереллиновой активности исследованных солей играют поляризуемость (α) и протоноакцепторная способность ($C_a^{\max}$) и через них – липофильность. В ряду производных N,N,N,-триэтилхолина также выявлен вклад протонодонорной способности (C_d). Стерический фактор (E_s) не играет большого значения. Возможная причина этого – линейное строение соединений, допускающее свободное вращение отдельных участков молекулы вокруг связей С-С и, тем самым, не препятствующее возникновению той конформации молекулы, которая необходима для ее взаимодействия с активным центром биомишени.

В отличие от производных холина и N,N,N,-триэтилхолина, для N,N-диалкилпиперидинийхлоридов, антигиббереллиновая активность и молекулярные параметры которых представлены в табл. 23, в проявлении активности существенную роль играют не только поляризуемость и протоноакцепторная способность, но и стерический фактор, причем наибольший вклад вносит именно он. Уравнение регрессии имеет вид:

$$\begin{aligned} \log(1/IC_{50}) = & 14.93(\pm 1.82) + 0.14(\pm 0.04) \alpha - 0.83(\pm 0.19) C_a^{\max} + \\ & 1.52(\pm 0.30) E_s \end{aligned} \quad (137)$$

$n=12; r^2=0.895; s=0.38; q^2=0.788$

Определяющее влияние стерического фактора связано, по-видимому, с тем, что четвертичный атом азота встроен в пиперидиновое кольцо. Это ограничивает конформационные возможности молекулы при ее взаимодействии с активным центром рецептора и предъявляет повышенные требования к молекулярным свойствам, которые привносят фрагменты R_1 и R_2 .

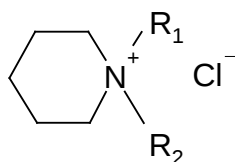
Таблица 22

Степень ингибирования биосинтеза гиббереллина (I, %), молекулярные параметры: поляризуемость α , Å³), максимальный свободноэнергетический протонноакцепторный дескриптор ($C_a^{\max}$), свободноэнергетический протоннодонорный дескриптор (C_d), стерический параметр (E_s) производных холина (I-VI) и N,N,N,-триэтилхолина (VII-XII)

№	Катион	Анион	I	α	$C_a^{\max}$	$-C_d$	$-E_s$
I	$[(CH_3)_3NCH_2CH_2OCH_3]^+$	Cl^-	19	14.16	1.46	0.00	6.55
II	$[(CH_3)_3NCH_2CH_2OC_2H_5]^+$	Cl^-	24	15.99	1.49	0.00	6.67
III	$[(CH_3)_3NCH_2CH_2OCH_2C_6H_5]^+$	Cl^-	89	23.81	1.25	0.00	7.10
IV	$[(CH_3)_3NCH_2CH_2Cl]^+$	Cl^-	85	13.62	0.39	0.00	6.48
V	$[(CH_3)_3NCH_2CH_2Br]^+$	Br^-	86	14.31	0.40	0.00	6.55
VI	$[(CH_3)_3NCH_2CH_2F]^+$	F^-	92	11.60	0.28	0.00	6.33
VII	$[(C_2H_5)_3NCH_2CH_2OH]^+$	Cl^-	16	17.83	1.54	1.93	8.04
VIII	$[(C_2H_5)_3NCH_2CH_2OCH_3]^+$	Cl^-	48	19.66	1.46	0.00	8.24
IX	$[(C_2H_5)_3NCH_2CH_2OC_2H_5]^+$	Cl^-	67	21.50	1.49	0.00	8.38
X	$[(C_2H_5)_3NCH_2CH_2OCH_2C_6H_5]^+$	Cl^-	85	29.32	1.49	0.00	8.91
XI	$[(C_2H_5)_3NCH_2CH_2Cl]^+$	Cl^-	69	19.12	0.39	0.00	8.11
XII	$[(C_2H_5)_3NCH_2CH_2Br]^+$	Br^-	76	19.82	0.40	0.00	8.19

Таблица 23

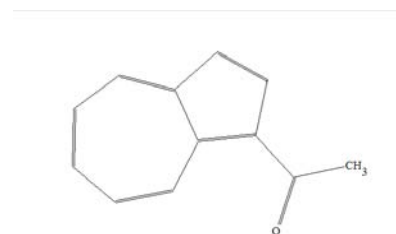
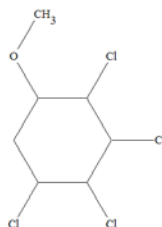
Антигипербереллиновая активность (IC_{50} , моль/л) и молекулярные параметры: поляризуемость α , ($\text{\AA}^3$), максимальный свободноэнергетический протоноакцепторный дескриптор ($C_a^{\max}$), стерический параметр (E_s) хлоридов N,N-диалкилпиперидиниев



№	R ₁	R ₂	log(1/IC ₅₀)	α	C _a ^{max}	-E _s
XIII	CH ₃	CH ₃	5.95	14.58	0.00	7.09
XIV	CH ₃	C ₄ H ₉	5.41	20.09	0.00	8.15
XV	C ₃ H ₇	C ₄ H ₉	4.28	23.76	0.00	9.10
XVI	CH ₃	CH ₂ CH ₂ OCH ₃	4.85	18.89	1.46	8.02
XVII	C ₂ H ₅	CH ₂ CH ₂ OCH ₃	3.62	20.72	1.46	8.64
XVIII	CH ₃	CH ₂ CH ₂ OC ₂ H ₅	4.15	20.72	1.49	8.19
XIX	C ₄ H ₉	CH ₂ CH ₂ OCH ₃	2.98	24.39	1.46	9.12
XX	CH ₂ C ₆ H ₅	CH ₂ CH ₂ OCH ₂ C ₆ H ₅	3.68	38.21	1.49	10.25
XXI	C ₃ H ₇	CH ₂ CH ₂ OCH ₃	3.00	22.56	1.46	8.92
XXII	CH ₂ C ₆ H ₅	CH ₂ CH ₂ OCH ₃	3.00	28.55	1.46	9.71
XXIII	CH ₃	CH ₂ CH ₂ OCH ₂ C ₆ H ₅	4.55	28.55	1.49	8.50
XXIV	CH ₃	C ₆ H ₅	5.25	22.53	0.32	8.66

Представленные в главе 4 локальные модели «структура – активность» могут быть использованы для предсказания биологической активности и для получения информации о структуре соединений с заданной величиной активности. Возьмем, например, модель (79). Оценка величины $\log(1/LC_{50})$ для планируемых к синтезу соединений не представляет сложности: достаточно рассчитать для 2D структуры величины $\sum C_a$ и подставить их в соответствующее уравнение. Разумеется, при этом должны быть соблюдены

все условия, касающиеся AD модели. Решение обратной задачи представляет известные трудности из-за ее некорректности (число неизвестных в модели превышает число уравнений). Один из вариантов ее решения может состоять в следующем. Предположим, нас интересуют структуры с величиной $\log(1/LC_{50})=1.00$ и интервалом изменения α на уровне $20 \div 22 \text{ \AA}^3$. Подставим соответствующие значения в уравнение (79) и определим интервал для C_a : $2.50 \div 3.09$. Далее, используя программу HYBOT (см. раздел 3.4), найдем структуры, соответствующие интервалам. В результате обнаружим 103 структуры, некоторые из которых приведены ниже.



Необходимо отметить, что модель (79) создана с использованием всего 150 соединений, при этом в интервал α $20 \div 22 \text{ \AA}^3$ попадает только восемь соединений. Решение обратной задачи с применением программы HYBOT позволяет выйти из круга структур обучающей выборки и значительно расширяет возможности для синтеза соединений с заданной величиной активности.

Выводы по главе

Разработанные на основе метода SS бинарные классификационные модели для соединений с МОА неполярный и полярный наркоз характеризуются высокой степенью правильного распознавания: 0.932 (обучающая выборка) и 0.914 (тестовая выборка) для пары неполярный наркоз/остальные МОА, 0.906 и 0.947 для пары полярный наркоз/остальные МОА и 0.952/1.000 для пары полярный/неполярный наркоз.

Установлено, что регрессионные QSAR модели острой токсичности органических соединений по отношению к рыбам *P. reticulata*, *P. promelas* и *O.*

mykiss могут быть сконструированы на основе двух физико-химических дескрипторов (молекулярной поляризуемости и Н-акцепторной способности) в случае неполярного наркоза и на основе трех физико-химических дескрипторов (молекулярной поляризуемости, Н-акцепторной и Н-донорной способности) в случае полярного наркоза.

Выявлено, что при анализе данных по острой токсичности 293 органических соединений по отношению к *P. reticulata* предложенный линейный дискриминантно – регрессионный подход позволил без использования априорных данных о типах токсического действия выделить группу химических соединений, состоящую, в основном, из соединений с МОА неполярный наркоз. Полученные регрессионные уравнения токсичности для этой группы близки к известным аналогичным моделям.

Установлено, что в случае производных холина и N,N,N-триэтилхолина, имеющих линейную структуру, наибольшее влияние на антигипбереллиновую активность оказывают поляризуемость, протоноакцепторный дескриптор и через них – липофильность соединений. Антигипбереллиновую активность пространственно более затрудненных солей N,N-диалкилпиперидиниев определяет, главным образом, стерический параметр, меньший вклад вносят поляризуемость и протоноакцепторный дескриптор.

Глава 5. Количественные глобальные модели «структура – биологическая активность» органических соединений

5.1. Модель АЛБТ

Одним из эффективных подходов, которые применяются при QSAR исследованиях в области токсикологии, в частности, при исследовании водной экотоксичности, является использование концепции типа действия (MOA) химических соединений. В самом простом варианте эта концепция предусматривает разделение молекул на две группы: неспецифического и специфического типа действия. В связи с этим вводится понятие «базовой» токсичности [289, 290] т.е. минимальной токсичности, присущей всем соединениям, и обусловленной неспецифическими взаимодействиями с биологическими структурами. Превышение порога базовой токсичности рассматривается как проявление специфических взаимодействий. Таким образом, основываясь на величине токсичности, все химические соединения можно разбить на две группы. К первой группе относятся молекулы, токсичность которых соответствует базовой токсичности, ко второй группе относятся все остальные молекулы.

Подходы, основанные на использовании классификации химических соединений с последующим построением регрессионных уравнений, хорошо зарекомендовали себя при конструировании моделей неспецифической токсичности соединений по отношению к водным организмам, например, к рыбам. Что касается других организмов, в частности, мышей и крыс, то создание подобных моделей неспецифической токсичности затруднено. Очевидно, это связано с большей сложностью этих биологических объектов. В связи с этим, можно отметить попытки использования концепции базовой токсичности для генерации количественных моделей токсичности

органических соединений при их ингаляционном [171] или пероральном [170] введении.

Количественное моделирование с использованием базовой токсичности обычно происходит с применением в качестве независимой переменной коэффициента распределения вещества в системе н-октанол – вода ($\log K_{ow}$). Получаемые при этом зависимости могут быть линейными и нелинейными. Ранее нами было установлено, что величина $\log K_{ow}$ может быть представлена в виде линейной комбинации двух дескрипторов: молекулярной поляризуемости α и суммы свободноэнергетических протоноакцепторных дескрипторов $\sum C_a$. Преимуществом этого подхода является использование в качестве независимых переменных расчетных величин, не требующих экспериментальных исследований. Кроме того, при увеличении размерности дескрипторного пространства возрастает вероятность правильной классификации.

При исследовании неспецифической острой токсичности химических соединений находят определенное применение гомологические ряды. В частности, наблюдаемые зависимости токсичности от $\log K_{ow}$ могут быть как линейного, так и нелинейного (параболического) типа [382, 383]. Следует отметить одну из особенностей гомологических рядов: они характеризуются приблизительно равными значениями $\sum C_a$. Таким образом, модели базовой токсичности гомологических рядов, где в качестве независимой переменной фигурирует $\log K_{ow}$, могут быть заменены моделями с использованием α . Учитывая, что доля гомологов в реальных выборках химических соединений невелика, с целью расширения круга рассматриваемых молекул, можно ввести понятие «псевдогомологов», т.е. соединений, у которых величины α варьируют, а $\sum C_a$ имеют приблизительно одинаковое значение. При использовании таких молекул для построения моделей базовой токсичности следует ожидать появления как линейных, так и нелинейных зависимостей токсичности от поляризуемости.

Алгоритм классификации с использованием анализа локальной базовой токсичности (АЛБТ) может быть описан в виде следующих шагов (рис. 14):

- 1) расчет величин молекулярной полярности $\sum C_a$ и свободнoэнергетических протоноакцепторных дескрипторов $\sum C_a$ и коэффициентов Танимото T_c для исходной выборки молекул;
- 2) деление исходной выборки на обучающую и тестовую;
- 3) задание шага изменения величин $\Delta \sum C_a$, $\Delta \alpha$ и токсичности $\Delta \log(1/LD_{50})$;
- 4) кластеризация молекул обучающей выборки с применением в качестве ведущей характеристики $\sum C_a$ (генерация интервалов с шагом $\Delta \sum C_a$ от 0 до $\max \sum C_a$);
- 5) разбиение каждого молекулярного кластера на ячейки с использованием α (интервалы с шагом $\Delta \alpha$ от 0 до $\max \alpha$);
- 6) обнаружение и идентификация в каждой ячейке молекулы с минимальным значением токсичности;
- 7) создание для каждого кластера ряда молекул с минимальными величинами токсичности;
- 8) расчет линейных или параболических регрессионных зависимостей токсичности от α для созданного ряда молекул;
- 9) расчет на основе полученных моделей величин базовой токсичности для всех молекул кластера;
- 10) классификация молекул кластера на две группы: (неполярный наркоз)/(все остальное) путем сравнения экспериментальных $(\log(1/LD_{50})_{\text{эксп}})$ и рассчитанных $(\log(1/LD_{50})_{\text{расч}})$ значений токсичности с учетом «коридора» базовой токсичности (интервалы $\log(1/LD_{50})_{\text{расч}} \pm \Delta \log(1/LD_{50})$ с шагом $\Delta \log(1/LD_{50})$ от 0 до $\max \Delta \log(1/LD_{50})$);
- 11) отнесение каждой молекулы обучающей выборки к соответствующему кластеру обучающей выборки на основе величины $\sum C_a$;

- 12) определение для каждой молекулы обучающей выборки ближайшего соседа с максимальной величиной T_c из соответствующего кластера;
- 13) классификация молекул обучающей выборки на две группы: (неполярный наркоз)/(все остальное) на основе принадлежности ближайшего соседа к той или иной группе;
- 14) расчет характеристик бинарной классификации молекул;
- 15) повторение шагов 11-14 для молекул тестовой выборки;
- 16) изменение величины $\Delta \sum C_a$, $\Delta \alpha$, $\Delta \log(1/LD_{50})$;
- 17) повторение шагов 3-16;
- 18) сортировка характеристик бинарной классификации молекул;
- 19) выбор лучшей модели.

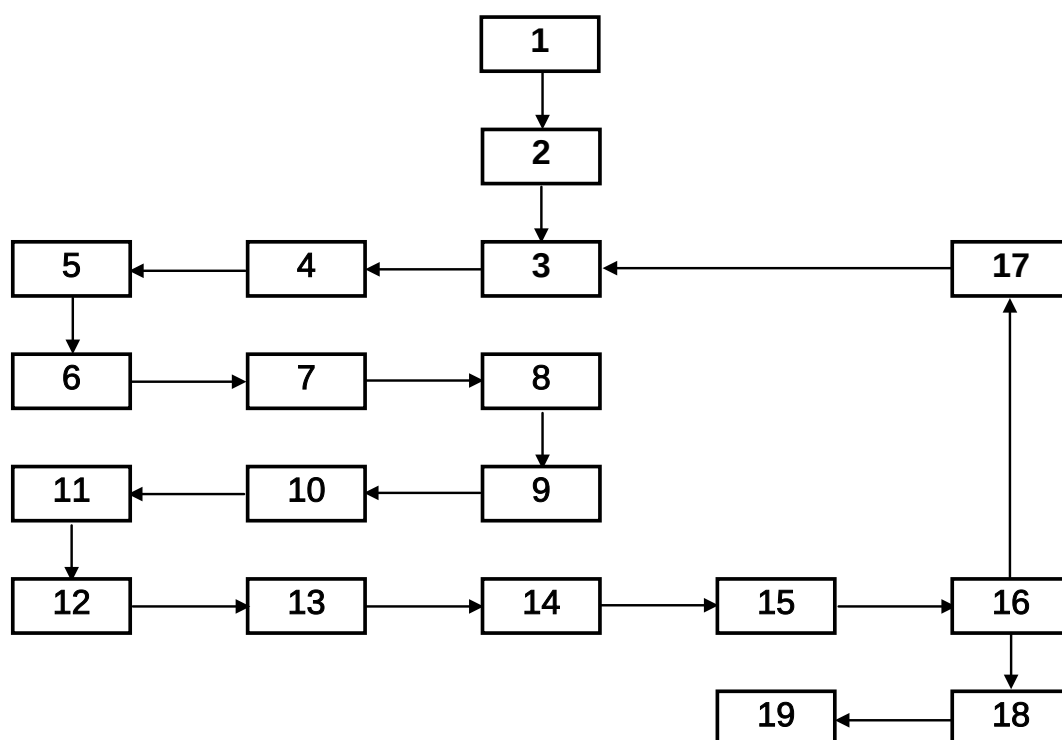


Рис. 14. Схема компьютерной программы АЛБТ

В настоящей работе в качестве сравнительных методов классификации химических соединений использованы такие методы, как LDA, KNN, SS, SVM, ANN. Эти методы требуют априорного знания о принадлежности соединений обучающей выборки к тому или иному классу. Так как для исследуемых 9618 органических соединений такая информация отсутствовала, то для их разделения на два класса (неполярный наркоз)/(все остальное) мы использовали метод аналогии, основываясь на данных по типам токсического действия органических соединений по отношению к водным организмам. Для этого были отобраны 42 классических инертных соединения, рассматриваемых как неполярные наркотики.

Таблица 24

Физико-химические дескрипторы: липофильность (MlogP), поляризуемость (α , Å³), сумма свободноэнергетических Н-акцепторных дескрипторов (ΣC_a) и базовая внутривенная токсичность (LD₅₀, ммоль/кг) органических соединений по отношению к мышам

№	Соединение	log(1/LD ₅₀)	MlogP	α	ΣC_a
1	1,2,4,5-Тетраметилбензол	-0.13	3.56	17.77	0.74
2	1-Бутанол	-0.71	0.80	8.75	1.60
3	1-Гексанол	0.00	1.59	12.42	1.59
4	1-Гептанол	0.10	1.94	14.26	1.65
5	1-Метилциклооктанол	0.40	2.20	17.15	1.60
6	1-Октанол	0.28	3.19	16.09	1.61
7	1-Пентанол	-0.32	1.21	10.59	1.61
8	1-Пропанол	-1.06	0.35	6.92	1.61
9	1-Пропил-1-циклопентанол	0.36	1.88	15.32	1.60
10	2-Бутанол	-1.01	0.80	8.75	1.62
11	2-Метил-1-пропанол	-0.75	0.80	8.75	1.57
12	2-Метил-2-бутанол	-0.84	1.21	10.59	1.63
13	2-Метил-2-нонанол	0.45	3.81	19.76	1.63
14	2-Метил-2-пентанол	-0.38	1.59	12.42	1.63
15	2-Метил-2-пропанол	-1.31	0.80	8.75	1.60
16	2-Метил-3-пентанол	-0.50	1.59	12.42	1.62
17	2-Октанол	0.30	2.27	16.09	1.60
18	2-Пропанол	-1.40	0.35	6.92	1.69
19	3-Метил-1-бутанол	-0.42	1.21	10.59	1.62
20	3-Метил-2-октанон	-0.10	2.45	17.44	1.96
21	3-Метил-3-гексанол	0.32	1.94	14.26	1.60
22	3-Метилциклопентадеканон	0.20	3.97	29.51	1.96
23	3-Пентанон	-0.77	1.06	10.10	1.88
24	3-Этил-2-пентанол	-0.19	1.94	14.26	1.60
25	4,4-Диметил-1-пентанол	-0.13	1.94	14.26	1.60
26	4-Метил-1-пентанол	-0.22	1.59	12.42	1.60
27	4-Метил-3-гептанол	-0.14	2.27	16.09	1.61
28	4-Пентилциклогексанол	0.48	2.79	20.82	1.62
29	4-Циклогексилметилциклогексанол	0.54	3.35	23.72	1.62
30	5-Нонанон	0.01	2.45	17.44	1.96
31	6-Ундеканон	0.16	3.04	21.11	1.96
32	8-Пентадеканон	-0.72	5.03	28.45	1.96

Продолжение табл. 24

№	Соединение	$\log(1/LD_{50})$	MlogP	α	ΣC_a
33	Дигептиловый эфир	-0.34	4.00	27.10	1.53
34	Диоктиловый эфир	-0.69	5.42	30.77	1.53
35	Дипентиловый эфир	-0.02	2.89	19.76	1.53
36	Дипропиловый эфир	-0.30	1.59	12.42	1.50
37	Дициклогексилкетон	0.54	3.20	23.24	1.90
38	Диэтиловый эфир	-1.13	0.80	8.75	1.54
39	Метанол	-2.17	-0.81	3.25	1.63
40	Циклогексанол	-0.43	1.20	11.65	1.67
41	Циклогексилбензол	0.38	3.74	20.67	0.67
42	Этанол	-1.63	-0.17	5.08	1.66

В табл. 24 представлены экспериментальные данные по острой внутривенной токсичности (LD_{50} , ммоль/кг, 24 час.) органических соединений с МОА неполярный наркоз по отношению к мышам, а также рассчитанные величины коэффициента распределения вещества в системе н-октанол-вода MlogP и HUBOT дескрипторов α ($\text{\AA}^3$), ΣC_a .

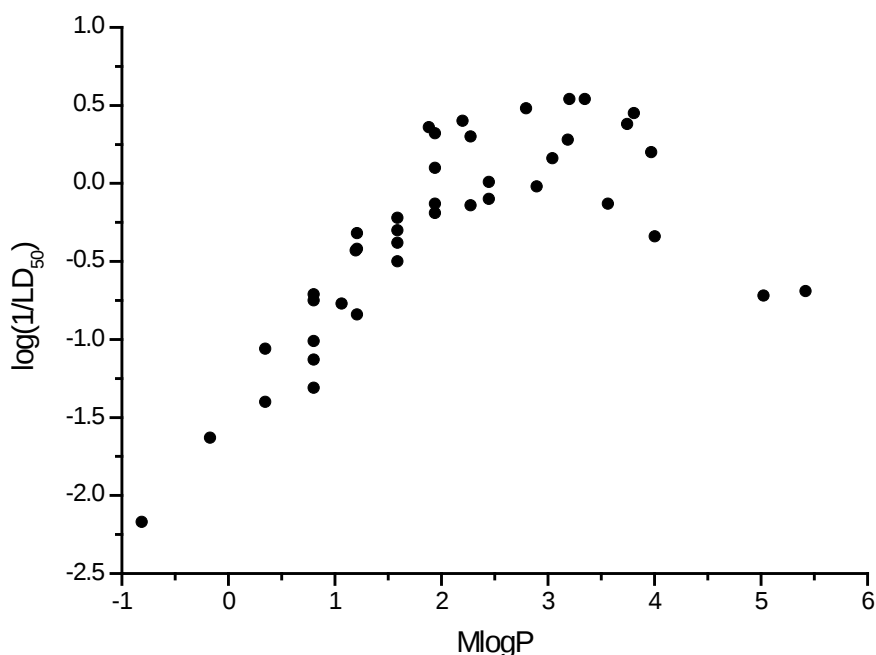


Рис. 15. Графическое представление зависимости базовой токсичности (LD_{50} , ммоль/кг) органических соединений от липофильности ($MlogP$)

На рис. 15 представлена графическая зависимость значений токсичности 42 органических соединений от рассчитанных значений коэффициента распределения в системе н-октанол-вода $MlogP$. Очевидно, что она имеет явный нелинейный характер. Применение параболической и билинейной моделей позволило получить следующие результаты:

$$\log(1/LD_{50}) = -1.56(\pm 0.10) + 1.12(\pm 0.08) MlogP - 0.18(\pm 0.02) (MlogP)^2 \quad (138)$$

$$n = 42; r^2 = 0.841; s = 0.26; F = 103; q^2 = 0.805$$

$$\log(1/LD_{50}) = -1.54(\pm 0.08) + 0.82(\pm 0.05) MlogP - 1.399(\pm 0.118) \log[0.001607 \cdot 10^{MlogP} + 1] \quad (139)$$

$$n = 42; r^2 = 0.868; s = 0.24; F = 129; q^2 = 0.852$$

Сравнение статистических критериев вышеуказанных моделей показывает преимущество билинейной модели. Применение NYBOT дескрипторов позволяет обсудить вклад водородной связи в неполярный наркоз.

$$\log(1/LD_{50}) = -2.89(\pm 1.56) + 0.29(\pm 0.08) \alpha - 0.009(\pm 0.001) \alpha^2 + 0.33(\pm 1.23) \Sigma C_a - 0.36(\pm 0.34) (\Sigma C_a)^2 + 0.04(\pm 0.05) \alpha \Sigma C_a \quad (140)$$

$n = 42; r^2 = 0.859; s = 0.25; F = 44.0; q^2 = 0.739$

Следует отметить, что ошибка определения коэффициента α превышает величину самого коэффициента. Поэтому можно упростить вышеприведенное уравнение и получить уравнение базовой токсичности только на основе молекулярной поляризуемости:

$$\log(1/LD_{50}) = -3.27(\pm 0.21) + 0.34(\pm 0.03) \alpha - 0.008(\pm 0.001) \alpha^2 \quad (141)$$

$n = 42; r^2 = 0.853; s = 0.25; F = 113; q^2 = 0.818$

Необходимо подчеркнуть, что в указанной выборке из 42 соединений представлены данные для мышей различного пола, возраста, условий содержания и эксперимента. Влияние этих факторов может давать различия в значениях $\log(1/LD_{50})$ на уровне ± 0.50 . Поэтому к 42 соединениям с базовой токсичностью мы решили добавить все соединения, которые имеют токсичность на уровне базовой (уравнение (139)) ± 0.50 . Графически это возможно описать с помощью «базового коридора» (рис. 16).

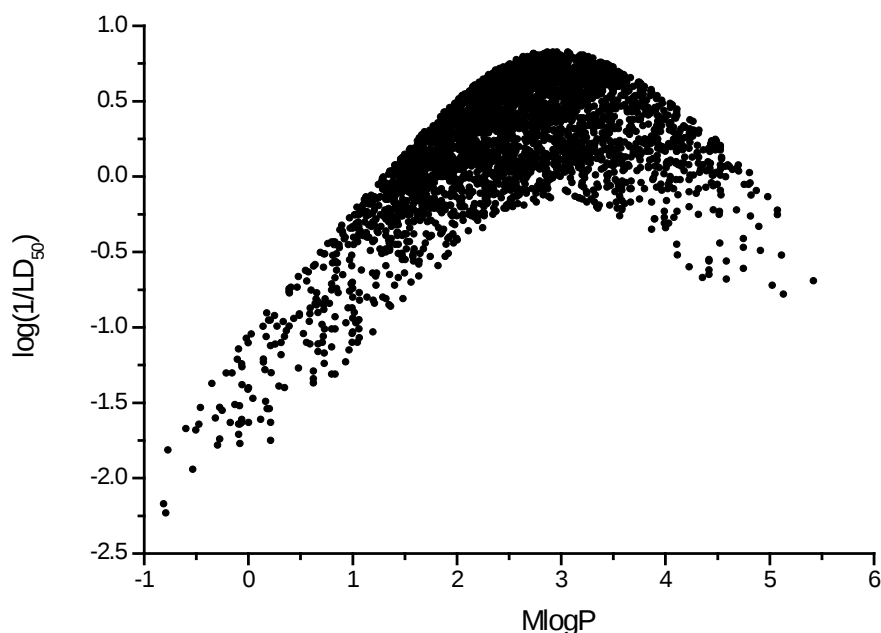


Рис. 16. Базовый коридор на основе уравнения (139)

Вошедшие в базовый коридор 3568 соединений имеют значения $MlogP$ в интервале $-0.81 \div 5.42$. Оставшиеся 6050 соединений расположены вне базового коридора: 5833 - над ним, 217 - под ним. Для всех соединений коридора была получена билинейная модель с хорошими статистическими параметрами:

$$\begin{aligned} \log(1/LD_{50}) = & -1.36(\pm 0.02) + 0.83(\pm 0.01) MlogP - \\ & 1.301(\pm 0.020) \log[0.002211 \cdot 10^{MlogP+1}] \end{aligned} \quad (142)$$

$n = 3568; r^2 = 0.678; s = 0.25; F = 3752; q^2 = 0.677$

Для оценки предсказательной способности полученной модели мы использовали кроме внутренней валидации (LOO), также внешнюю валидацию. Для этого 3568 соединений базового коридора были разбиты на обучающую (2855 соединений) и тестовую (713 соединений) выборки. В результате для обучающей выборки было получено следующее уравнение:

$$\log(1/LD_{50}) = -1.36(\pm 0.02) + 0.84(\pm 0.01) M\log P - 1.310(\pm 0.020) \log[0.002256 P + 1] \quad (143)$$

$$n = 2855; r^2 = 0.674; s = 0.25; F = 2948; q^2 = 0.673$$

Предсказание токсичности для тестового ряда на основе уравнения (143) дало следующие результаты:

$$\log(1/LD_{50})_{\text{эксп}} = -0.01(\pm 0.01) + 0.99(\pm 0.02) \log(1/LD_{50})_{\text{расч}} \quad (144)$$

$$n = 713; r^2 = 0.695; s = 0.24; F = 1617; q^2 = 0.693$$

Статистические критерии последнего уравнения показывают хорошую стабильность QSAR моделей базовой токсичности.

В табл. 25 представлены соединения различных классов в базовом коридоре и областях с большими и меньшими значениями токсичности.

Таблица 25

Распределение 9618 органических соединений по функциональным группам и
МОА

Химическая группа		МОА		
		Базовый наркоз (3568, 37%)	Специфические (5833, 61%)	Метаболиты (217, 2%)
R-O-H	3420	1191	2056	173
Ph-O-H	887	259 (29%)	617 (70%)	11 (1%)
C-O-H	2534	932 (37%)	1440 (57%)	162 (6%)
R-C≡N	314	77	234	3
Ph-C≡N	50	19 (38%)	31 (62%)	0 (0%)
C-C≡N	264	58 (22%)	203 (77%)	3 (1%)
R-CF₃	264	78	182	4
Ph-CF ₃	52	25 (48%)	27 (52%)	0 (0%)
C-CF ₃	212	53 (25%)	155 (73%)	4 (2%)
R-C(NH₂)=O	2376	858	1376	142
Ph-C(NH ₂)=O	388	154 (40%)	167 (43%)	67 (17%)
C-C(NH ₂)=O	1988	704 (35%)	1209 (61%)	75 (4%)
R-SO₂	315	121	191	3
Ph-SO ₂	243	106 (44%)	134 (55%)	3 (1%)
C-SO ₂	72	15 (21%)	57 (79%)	0 (0%)
R-S-H	63	28	35	0
Ph-S-H	4	3 (75%)	1 (25%)	0 (0%)
C-S-H	59	25 (42%)	34 (58%)	0 (0%)
R-S(=O)-R	111	45	66	1
Ph-S(=O)-Ph	23	13 (57%)	10 (43%)	0 (0%)
Ph-S(=O)-C	31	15 (48%)	16 (52%)	0 (0%)
C-S(=O)-C	58	17 (29%)	40 (69%)	1 (2%)
R-NO₂	389	154	233	2
Ph-NO ₂	322	133 (41%)	188 (58%)	1 (1%)
C-NO ₂	67	21 (31%)	45 (67%)	1 (2%)
R-Cl	1388	597	785	6
Ph-Cl	827	390 (47%)	432 (52%)	5 (1%)
C-Cl	561	207 (36.5%)	353 (63%)	1 (0.5%)
R-Br	302	121	178	3
Ph-Br	135	62 (46%)	70 (52%)	3 (2%)
C-Br	167	59 (35%)	108 (65%)	0 (0%)
R-I	367	166	93	108
Ph-I	316	148 (47%)	64 (20%)	104 (33%)

Продолжение табл. 25

Химическая группа		МОА		
		Базовый наркот	Специфические (5833, 61%)	Метаболиты (217, 2%)
C-I	51	18 (35%)	29 (57%)	4 (8%)
R-F	521	164	349	8
Ph-F	171	65 (38%)	103 (60%)	3 (2%)
C-F	350	99 (28%)	246 (70%)	5 (2%)
R-C(=O)-O-H	920	456	306	158
Ph-C(=O)-O-H	216	120 (56%)	35 (16%)	61 (28%)
C-C(=O)-O-H	704	336 (48%)	271 (38%)	97 (14%)
R-C(=O)-H	76	21	55	0
Ph-C(=O)-H	22	13 (59%)	9 (41%)	0 (0%)
C-C(=O)-H	54	8 (15%)	46 (85%)	0 (0%)
R-C(=O)-R	1071	254	805	12
Ph-C(=O)-Ph	76	18 (24%)	57 (75%)	1 (1%)
Ph-C(=O)-C	251	82 (33%)	167 (66.5%)	2 (0.5%)
C-C(=O)-C	744	154 (21%)	581 (78%)	9 (1%)
R-S-R	826	363	452	11
C-S-C	574	240 (42%)	323 (56%)	11 (2%)
Ph-S-C	124	69 (56%)	55 (44%)	0 (0%)
Ph-S-Ph	128	54 (42%)	74 (68%)	0 (0%)
R-O-R	4042	1178	2804	60
C-O-C	2419	728 (30%)	1647 (68%)	44 (2%)
Ph-O-C	1579	436 (28%)	1128 (71%)	15 (1%)
Ph-O-Ph	44	14 (32%)	29 (66%)	1 (2%)
R-NH₂	725	257	462	6
Ph-NH ₂	213	110 (52%)	99 (46%)	4 (2%)
C-NH ₂	512	147 (29%)	363 (70.5%)	2 (0.5%)
R-NH-R	839	306	533	6
Ph-NH-C	282	149 (53%)	133 (47%)	0 (0%)
Ph-NH-Ph	31	27 (87%)	4 (13%)	0 (0%)
C-NH-C	532	130 (24%)	396 (75%)	6 (1%)
R-SO₂(NH₂)	183	69	111	3
Ph-SO ₂ (NH ₂)	153	62 (41%)	88 (58%)	3 (1%)
C-SO ₂ (NH ₂)	30	7 (23%)	23 (77%)	0 (0%)
R-N(R)-R	1969	715	1235	19
C-N(C)-C	1634	580 (35%)	1036 (63%)	18 (2%)
Ph-N(C)-C	233	95 (41%)	137 (58.6%)	1 (0.4%)
Ph-N(Ph)-C	101	39 (39%)	62 (61%)	0 (0%)
Ph-N(Ph)-Ph	1	1 (100%)	0 (0%)	0 (0%)

Продолжение табл. 25

Химическая группа		МОА		
		Базовый наркоз (3568, 37%)	Специфические (5833, 61%)	Метаболиты (217, 2%)
R-C(=O)-O-R	888	172	707	9
C-C(=O)-O-Ph	41	19 (46%)	20 (63.5%)	2 (0.5%)
C-C(=O)-O-C	635	109 (17%)	520 (82%)	6 (1%)
Ph-C(=O)-O-C	202	40 (20%)	161 (79.7%)	1 (0.3%)
Ph-C(=O)-O-Ph	10	4 (40%)	6 (60%)	0 (0%)
P=O	240	99 (41%)	135 (56%)	6 (3%)
R-P(=O)(R)-R	167	61	103	3
C-P(=O)(C)-O	31	15 (48.4%)	15 (48.4%)	1 (3.2%)
C-P(=O)(O)-O	52	17 (33%)	33 (63%)	2 (4%)
O-P(=O)(O)-O	63	18 (29%)	45 (71%)	0 (0%)
C-P(=O)(C)-C	21	11 (52%)	10 (48%)	0 (0%)
R-Pyridine	89	27	62	0
C-Pyridine	86	25 (29%)	61 (71%)	0 (0%)
Ph-Pyridine	3	2 (67%)	1 (33%)	0 (0%)

Дифференциация всех исследуемых соединений показывает, что 37% из них относится к инертным, 61% более токсичны и находятся выше базового коридора (полярный наркоз, реактивные, специфические), 2% соединений менее токсичны и расположены ниже коридора. Для начала можно отметить существенное влияние фенильного кольца на токсичность различных функциональных групп. В случае, если одним из заместителей (R) является фенильная группировка, происходит увеличение числа инертных соединений, по сравнению с алифатическими заместителями, для R-C≡N, R-CF₃, R-C(NH₂)=O, R-SO₂, R-S-H, R-S(=O)-R, R-NO₂, R-Cl, R-Br, R-I, R-F, R-C(=O)-O-H, R-C(=O)-H, R-S-C, R-NH₂, R-NH-R, R-SO₂(NH₂), R-C(=O)-O-C. В случае R-O-H и R-C(=O)-C наблюдается обратная зависимость.

Рассмотрение различных классов химических соединений показывает, что приблизительно то же самое соотношение наблюдается для соединений со следующими группами: C-OH, Ph-C≡N, C-C(NH₂)=O, Ph-NO₂, C-Cl, C-Br, C-I, C-F, Ph-F, C-N(C)-C, Ph-N(Ph)-C, C-P(=O)(O)-O. Таким образом, можно заключить, что нет какого-либо существенного влияния данных групп на

другие функциональные группировки. В случае Ph-O-H, C-C $\equiv$ N, C-CF₃, C-SO₂, C-F, C-C(=O)-H, Ph-C(=O)-Ph, C-C(=O)-C, Ph-O-C, C-NH₂, C-NH-C, C-SO₂(NH₂), C-C(=O)-O-C, Ph-C(=O)-O-C, O-P(=O)(O)-O, C-Pyridine число более токсичных соединений увеличивается.

В табл. 26 представлены результаты бинарной классификации 9618 органических соединений, обладающих острой внутривенной токсичностью по отношению к мышам, с использованием различных методов в пространстве из 16 HYBOT дескрипторов (селекция дескрипторов при $r=0.9$) (LDA, KNN, SVM, ANN), фрагментных дескрипторов (SS), $\alpha \sum C_a$ дескрипторов (АЛБТ).

Следует отметить, что среди всех использованных методов самые худшие результаты показывает метод LDA, а самые лучшие – АЛБТ. Так, в случае LDA, величина ACC (интегральный показатель качества классификации), составляет 0.666 (обучающая выборка) и 0.652 (тестовая выборка).

Таблица 26

Чувствительность (SN), специфичность (SP) и точность (ACC) при бинарной классификации химических соединений с МОА неполярный наркоз (класс 1) и другие МОА (класс 2)

№	Метод	Обучающая выборка				Тестовая выборка			
		n_1/n_2	SN	SP	ACC	n_1/n_2	SN	SP	ACC
1	LDA	2811/4884	0.291	0.881	0.666	757/1166	0.284	0.890	0.652
2	KNN_7	2811/4884	0.605	0.807	0.733	757/1166	0.594	0.813	0.727
3	SS_5	2811/4884	0.587	0.798	0.721	757/1166	0.568	0.810	0.715
4	SVM	2811/4884	0.855	0.947	0.913	757/1166	0.567	0.852	0.739
5	ANN	2811/4884	0.406	0.853	0.690	757/1166	0.394	0.846	0.668
6	АЛБТ	1959/5715	0.992	1.000	0.998	509/1405	0.650	0.887	0.824

Примечание: n_1 – число соединений в первом классе, n_2 – число соединений во втором классе.

При использовании метода локальной базовой токсичности ACC достигает 0.998 и 0.824 соответственно. При этом величина (SN) правильного распознавания молекул первой группы (неполярный наркоз) для обучающей и тестовой выборки составляет 0.992 и 0.650, а величина (SP) правильной классификации молекул с другими МОА находится на уровне 1.000 и 0.887 соответственно. Подходы, основанные на анализе ближайшего соседа (KNN_7 и SS_5), показывают промежуточные результаты. Величина ACC колеблется от 0.715 до 0.757 в тестовой выборке. Общей чертой для всех методов классификации является то, что величина чувствительности (SN) во всех случаях меньше специфичности (SP), т.е. вероятность правильного распознавания молекул с МОА неполярный наркоз ниже, чем вероятность правильного распознавания молекул с другими МОА. Вероятно, это связано с наличием в исследуемых данных системных погрешностей, обусловленных

ошибками в определении величин токсичности и структурным несоответствием из-за присутствия метаболитов.

5.2. Модель AMP

Острая токсичность химических соединений по отношению к млекопитающим является сложным и плохо изученным феноменом, состоящим из множественных неспецифичных взаимодействий соединений с биологическими мембранами, а также специфического связывания с большим числом биологических мишеней внутри организма. Как результат этой сложности и плохого механистического понимания, имеются трудности в разработке QSAR моделей острой токсичности по отношению к млекопитающим, основанных на механизмах действия, и число QSAR исследований в этой области мало [9]. Недавний детальный анализ 150 опубликованных QSAR моделей острой токсичности по отношению к мышам и крысам привел к заключению о том, что многие из этих моделей имели ограниченную пользу для предсказания величин LD₅₀ из-за их плохого или ограниченного статистического качества, или потому что они были получены на слишком ограниченных рядах данных [7]. Судя по литературным данным, в большинстве исследований устанавливается связь между величинами острой токсичности и липофильности соединений, хотя общепризнанно, что описание такого сложного феномена как токсичность с помощью только этого, связанного с молекулярным транспортом, параметра недостаточно [268, 387]. В связи с этим следует отметить работу [297], в которой получены консенсусные модели острой токсичности соединений по отношению к крысам.

Также необходимо отметить, что имеются доказательства того, что общие QSAR регрессионные уравнения для какого-либо свойства (активности) разнообразных соединений являются аппроксимацией, которая приглаживает реальный профиль свойства (активности). [388]. Поэтому возможно, что на

такой морщинистой поверхности могут существовать как «вершины активности», в результате совершенного молекулярного распознавания биологической мишени, так и «впадины активности», обусловленные отсутствием активности из-за метаболизма или других реакций. В этих условиях маловероятно ожидать, что традиционный QSAR будет эффективным в улавливании реальной активности разнообразных химических соединений.

В качестве средства преодоления этой проблемы, в литературе [180] предложен простой подход, основанный на использовании арифметической средней токсичности ближайших структурных соседей. В настоящей работе это направление получило дальнейшее развитие, связанное не только с анализом ближайших структурных соседей, но и определением физико-химического сходства молекул. Модель AMP (арифметическое среднее свойство) была использована для предсказания острой токсичности химических соединений по отношению к мышам при внутривенном введении. Исходный массив данных содержал величины LD_{50} (ммоль/кг, 24 час.) для 10241 органического соединения.

Алгоритм модели AMP может быть описан в виде четырех шагов.

Шаг 1. Для каждой анализируемой молекулы определялась величина структурного сходства (в виде коэффициента Танимото T_c) с остальными молекулами. Величины T_c упорядочивались по убыванию. Путем задания минимального граничного значения T_c формировались различного размера первичные кластеры ближайших структурных соседей.

Шаг 2. Для всех молекул рассчитывался ряд дескрипторов.

Шаг 3. С использованием каждого дескриптора в отдельности и первичного структурного кластера для каждой молекулы рассчитывался вторичный кластер. Для этого из молекул первичного кластера выбирались такие пары ближайших соседей, которые имели значение дескриптора больше и меньше, чем у анализируемой молекулы. Размер вторичного кластера регулировался путем задания числа молекулярных пар.

Шаг 4. Токсичность анализируемой молекулы рассчитывалась как арифметическое среднее токсичностей молекулярных пар.

Схематическое объяснение модели АМР представлено на рис. 17.

В этом примере молекулярная поляризуемость (α) используется как дескриптор для формирования вторичного кластера соседей на основе двух молекулярных пар. Анализируемое соединение ($\log(1/LD_{50})=1.96$) имеет 10 ближайших структурных соседей (первичный кластер). По сравнению с анализируемым соединением, 8 соседей имеют меньшую величину (α (1 группа), 2 соседа – большую величину α (2 группа). Для образования первой пары берется сосед с максимальным значением T_c ($\log(1/LD_{50})=2.46$, 1 группа). Ближайшим соседом из 2 группы оказывается соединение с $\log(1/LD_{50})=3.76$. Для формирования второй пары берется еще неиспользованное соединение с максимальным значением T_c ($\log(1/LD_{50})=0.71$, 2 группа). Ближайшим соседом из 1 группы оказывается соединение с $\log(1/LD_{50})=0.30$. На основе двух выбранных молекулярных пар можно оценить величину токсичности для анализируемого соединения: $\log(1/LD_{50})=(2.46+3.76+0.71+0.30)/4=1.81$.

Кроме α , в работе использовались и другие HUBOT дескрипторы. Наилучшие результаты были получены на основе композитного дескриптора $\sum C_{ad}/\alpha$. Поэтому этот дескриптор был использован, чтобы получить результаты, описанные ниже.

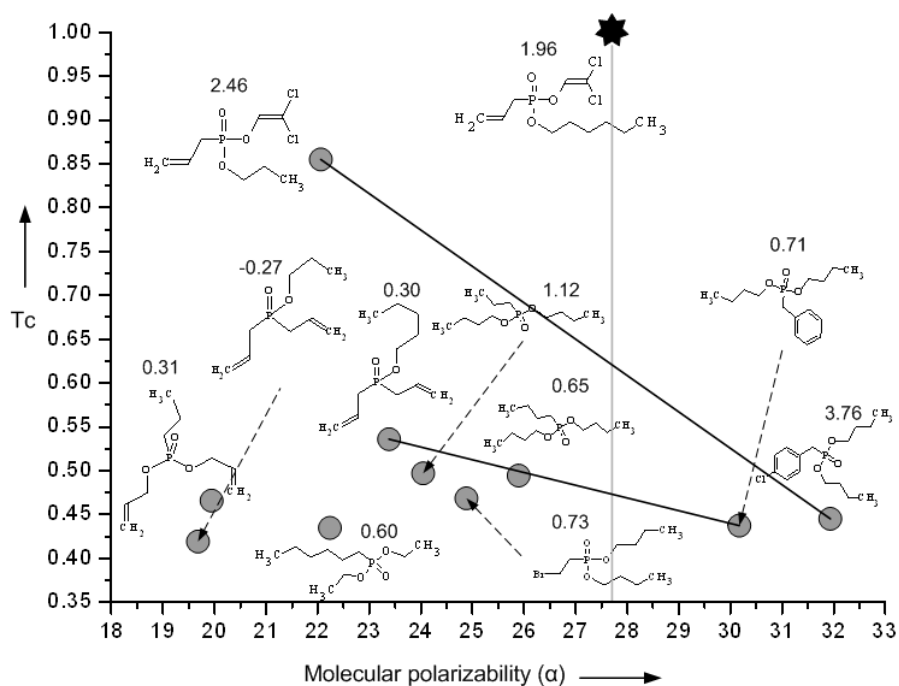


Рис. 17. Схема выбора соседей в методе AMP. * - анализируемое соединение, • - соседи

Профиль токсичности анализируемых данных представлен на рис. 18 (а). Полный интервал величин $\log(1/LD_{50})$ для 10241 соединений составляет 7.49 единиц (от -3.14 до 4.35).

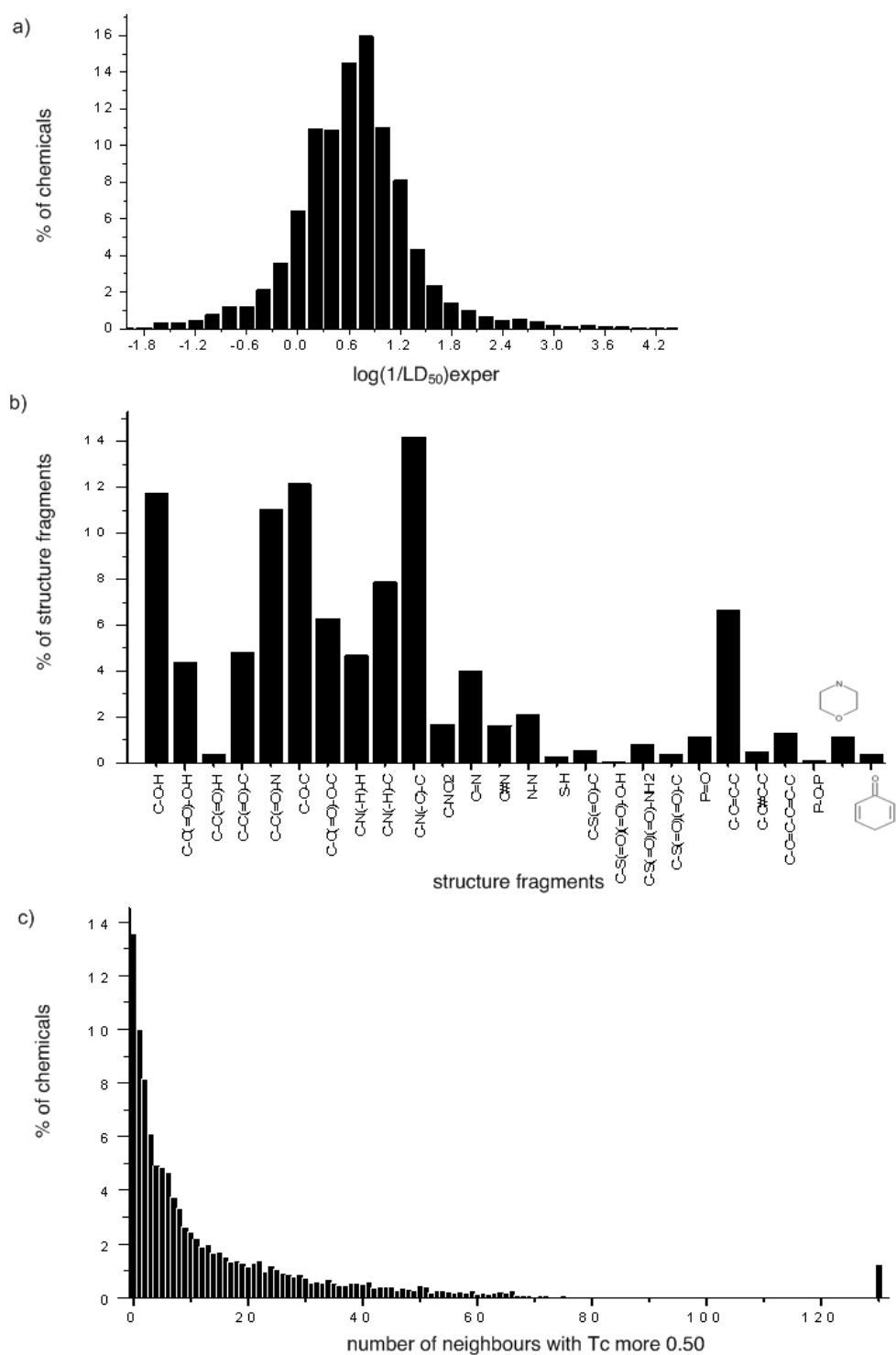


Рис. 18. (a) Профиль токсичности исследуемого ряда из 10241 соединения; (b) профиль функциональных химических групп; (c) гистограмма распределения числа структурных соседей с $T_c \geq 0.50$

Однако, около 97% всех соединений имеют величины $\log(1/LD_{50})$ в интервале от -1.50 до 2.50. Размер этого интервала и нормальное распределение соединений являются адекватными для построения модели.

Рис. 18 (b) иллюстрирует химическое разнообразие исследуемого ряда соединений. Левая часть рисунка идентифицирует присутствие ключевых функциональных химических групп и демонстрирует, что ряд данных содержит значимое число соединений в различных классах.

Рис. 18 (c) обеспечивает информацией о числе соединений, имеющих структурных соседей с $T_c \geq 0.50$. Было установлено, что около 13% соединений не имеют таких структурных соседей вообще, около 10% имеют только одного такого соседа, 8% имеют только два структурных соседа, 6% соединений имеют три соседа и около 5% имеют четыре структурных соседа с $T_c \geq 0.50$. Это означает, что концепция структурного сходства должна осторожно применяться к этим 42% анализируемого ряда соединений, и что ошибка определения токсичности таких соединений должна быть выше по сравнению с соединениями, имеющими больше структурных соседей.

Табл. 27 содержит статистические характеристики и параметры уравнения (145) с различными граничными значениями структурного сходства (T_c) и различным числом ближайших соседних пар (NP):

$$\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}} \quad (145)$$

Таблица 27

Параметры и статистические характеристики уравнения $\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ при различных значениях структурного сходства (T_c) и различном числе ближайших соседних пар (NP)

T_c	NP	a_0	a_1	n	r^2	s	q^2	F
≥ 0.00	1	0.15(±0.01)	0.77(±0.01)	10228	0.456	0.52	0.456	8521
	2	0.09(±0.01)	0.88(±0.01)	10227	0.484	0.51	0.484	9604
	3	0.07(±0.01)	0.94(±0.01)	10226	0.484	0.51	0.482	9572
	4	0.05(±0.01)	0.97(±0.01)	10225	0.477	0.51	0.475	9307
	5	0.04(±0.01)	0.99(±0.01)	10224	0.470	0.51	0.469	9048
≥ 0.40	1	0.12(±0.01)	0.81(±0.01)	8256	0.533	0.46	0.533	9416
	2	0.06(±0.01)	0.92(±0.01)	7085	0.577	0.43	0.577	9672
	3	0.02(±0.01)	0.98(±0.01)	6213	0.596	0.42	0.593	9152
	4	0.00(±0.01)	1.02(±0.01)	5534	0.610	0.41	0.609	8663
	5	0.00(±0.01)	1.03(±0.01)	4993	0.605	0.41	0.605	7599
≥ 0.50	1	0.09(±0.01)	0.86(±0.01)	6241	0.593	0.44	0.593	8960
	2	0.03(±0.01)	0.96(±0.01)	4624	0.643	0.40	0.642	8315
	3	0.00(±0.01)	1.01(±0.01)	3655	0.660	0.39	0.657	7082
	4	-0.01(±0.01)	1.04(±0.01)	2983	0.670	0.39	0.665	6049
	5	-0.02(±0.01)	1.04(±0.01)	2496	0.681	0.37	0.681	5306
≥ 0.60	1	0.08(±0.01)	0.88(±0.01)	4177	0.653	0.41	0.653	7795
	2	0.03(±0.01)	0.98(±0.01)	2614	0.690	0.37	0.690	5823
	3	0.00(±0.01)	1.02(±0.02)	1834	0.699	0.37	0.696	4264
	4	-0.01(±0.02)	1.04(±0.02)	1343	0.708	0.36	0.700	3253
	5	-0.01(±0.02)	1.05(±0.02)	1012	0.710	0.35	0.705	2469
≥ 0.70	1	0.07(±0.01)	0.91(±0.01)	2557	0.692	0.39	0.692	5656
	2	0.04(±0.02)	0.98(±0.02)	1323	0.706	0.36	0.705	3171
	3	0.02(±0.02)	1.02(±0.03)	783	0.718	0.37	0.711	1984
	4	-0.02(±0.02)	1.10(±0.03)	522	0.783	0.33	0.773	1871
≥ 0.80	1	0.05(±0.01)	0.94(±0.02)	1153	0.760	0.34	0.760	3648
	2	0.03(±0.02)	1.00(±0.03)	491	0.762	0.33	0.756	1567

Очевидно, что при возрастании границы структурного сходства и числа пар, свободный член уравнения приближается к 0, а угловой коэффициент приближается к 1. В то же самое время, коэффициент линейной корреляции возрастает, а стандартное отклонение убывает. Однако, этот процесс сопровождается уменьшением числа соединений, вовлеченных в расчет.

В принципе, почти для всех химических соединений изучаемого ряда могут быть рассчитаны величины токсичности на основе модели АМР с использованием $T_c \geq 0.00$ и $NP=2$. Только 14 соединений из 10241 не имеют двух пар соседей с большим и меньшим значением выбранного дескриптора ($\sum C_{ad}/\alpha$). Параметры уравнения (145) в этом случае удовлетворительные и стандартное отклонение сравнимо с ожидаемыми экспериментальными ошибками измерения токсичности.

Тем не менее, статистические характеристики уравнения (145), являясь средними величинами, не отражают специфических отклонений и ошибок оценивания из-за различных уровней структурного сходства. По этой причине мы предложили в качестве альтернативы каскадный расчет токсичности, в котором АМР модели появляются последовательно.

На самом высоком уровне А граница структурного сходства $T_c \geq 0.80$ и число вовлеченных в расчет структурных пар $NP=2$. Этому уровню соответствует 491 соединение (4.79% от общего числа соединений). Результат применения этой модели может быть представлен в следующем виде:

$$\log(1/LD_{50})_{\text{эксп}} = 0.03(\pm 0.02) + 1.00(\pm 0.03) \log(1/LD_{50})_{\text{расч}} \quad (146)$$

$$n=491; r^2=0.762; s=0.33; F=1567; q^2=0.756$$

Таким образом, токсичность этих соединений может быть оценена со стандартным отклонением $s=0.33$. Графическое сравнение экспериментальных и рассчитанных величин токсичности соединений на уровне А представлено на рис. 19.

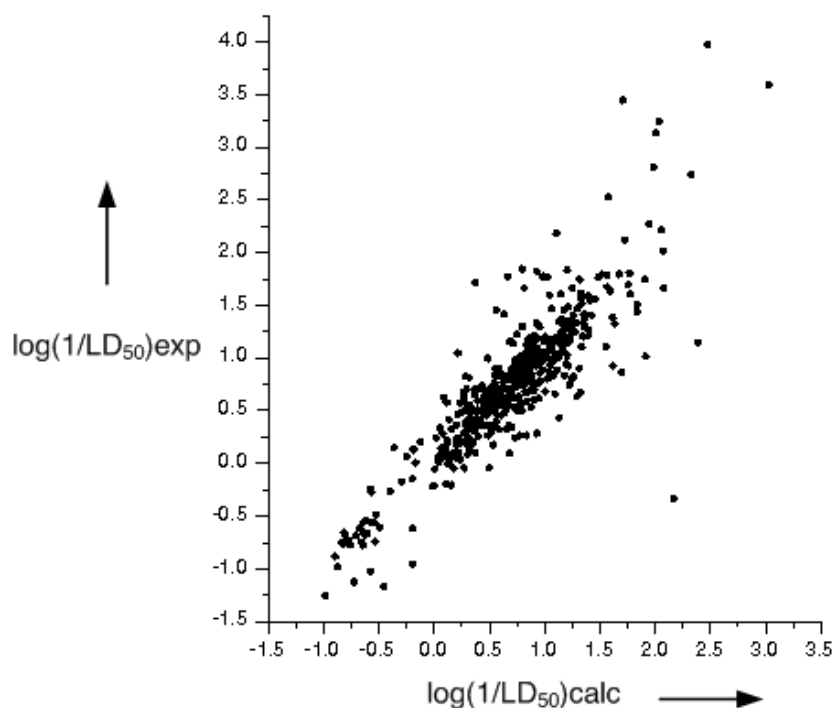


Рис. 19. Графическое сравнение экспериментальных и рассчитанных на основе модели АМР (уровень А) величин острой токсичности (LD_{50} , ммоль/кг)

Следующий уровень В включает в себя соединения с $T_c \geq 0.70$ и $NP=2$, исключая соединения уровня А. Для таких 832 соединений (8.12% от общего числа соединений) стандартное отклонение равно 0.39:

$$\log(1/LD_{50})_{\text{эксп}} = 0.04(\pm 0.02) + 0.96(\pm 0.02) \log(1/LD_{50})_{\text{расч}} \quad (147)$$

$$n=832; r^2=0.673; s=0.39; F=1708; q^2=0.671$$

Следующий уровень С включает в себя соединения с $T_c \geq 0.60$ и $NP=2$, исключая соединения уровня А и В. Для этих 1291 соединения (12.60% от общего числа соединений) стандартное отклонение равно 0.38:

$$\log(1/LD_{50})_{\text{эксп}} = 0.01(\pm 0.02) + 0.97(\pm 0.02) \log(1/LD_{50})_{\text{расч}} \quad (148)$$

$$n=1291; r^2=0.673; s=0.38; F=2654; q^2=0.671$$

Было установлено, что 2010 соединений (19.63% от общего числа соединений) соответствует уровню D, имеющему $T_c \geq 0.50$ и NP=2, исключая соединения уровня A, B и C:

$$\log(1/LD_{50})_{\text{эксп}} = 0.04(\pm 0.01) + 0.94(\pm 0.02) \log(1/LD_{50})_{\text{расч}} \quad (149)$$

$n=2010$; $r^2=0.581$; $s=0.44$; $F=2785$; $q^2=0.580$

2461 соединение (24.03%) с $T_c \geq 0.40$ и NP=2, исключая соединения уровня A, B, C и D, были включены в уровень E. Соответствующее уравнение имеет недостаточный угловой коэффициент и стандартное отклонение, приближающееся по своей величине к ошибке измерения токсичности:

$$\log(1/LD_{50})_{\text{эксп}} = 0.10(\pm 0.01) + 0.82(\pm 0.02) \log(1/LD_{50})_{\text{расч}} \quad (150)$$

$n=2461$; $r^2=0.442$; $s=0.48$; $F=1947$; $q^2=0.440$

Оставшиеся 3142 соединения (30.83%) не имеют хороших структурных соседей. Однако, для них также были рассчитаны токсичности с использованием модели AMP с двумя парами структурных соседей (уровень F). Параметры и статистические характеристики следующего уравнения свидетельствуют о низкой точности расчета токсичности для таких соединений:

$$\log(1/LD_{50})_{\text{эксп}} = 0.16(\pm 0.01) + 0.83(\pm 0.02) \log(1/LD_{50})_{\text{расч}} \quad (151)$$

$n=3142$; $r^2=0.337$; $s=0.64$; $F=1593$; $q^2=0.335$

Таким образом, использование каскадной модели AMP позволило оценить острую токсичность 7085 соединений (около 69% от анализируемого ряда данных) по отношению к мышам со стандартным отклонением, сравнимым с экспериментальной ошибкой определения. Для оставшихся 3142

соединений стандартное отклонение равно 0.64, что эквивалентно приблизительно одной трети интервала токсичности для 97% изученных соединений. Поэтому это лучше, чем предсказание токсичности таких соединений качественным образом (низкая, средняя или высокая токсичность).

Для валидации модели мы разделили исследуемый ряд из 10241 соединения на обучающую (8193 соединения) и тестовую (2048 соединений) выборки. Токсичность соединений тестового ряда была предсказана с использованием только соединений обучающего ряда.

Таблица 28

Параметры и статистические характеристики уравнения $\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ для обучающей и тестовой выборок при различных уровнях структурного сходства

Уровень	Обучающая выборка (8193 соединения)	Тестовая выборка (2048 соединений)
A	$\log(1/LD_{50})_{\text{эксп}} = 0.02(\pm 0.03) + 1.02(\pm 0.03) \log(1/LD_{50})_{\text{расч}}$ $n = 317, r^2 = 0.735, s = 0.35, q^2 = 0.729, F = 876$	$\log(1/LD_{50})_{\text{эксп}} = 0.06(\pm 0.04) + 0.91(\pm 0.05) \log(1/LD_{50})_{\text{расч}}$ $n = 81, r^2 = 0.796, s = 0.26, q^2 = 0.780, F = 307$
B	$\log(1/LD_{50})_{\text{эксп}} = 0.07(\pm 0.03) + 0.97(\pm 0.03) \log(1/LD_{50})_{\text{расч}}$ $n = 513, r^2 = 0.648, s = 0.40, q^2 = 0.645, F = 940$	$\log(1/LD_{50})_{\text{эксп}} = 0.06(\pm 0.05) + 0.92(\pm 0.05) \log(1/LD_{50})_{\text{расч}}$ $n = 125, r^2 = 0.707, s = 0.40, q^2 = 0.692, F = 297$
C	$\log(1/LD_{50})_{\text{эксп}} = 0.04(\pm 0.02) + 0.90(\pm 0.02) \log(1/LD_{50})_{\text{расч}}$ $n = 950, r^2 = 0.635, s = 0.37, q^2 = 0.633, F = 1652$	$\log(1/LD_{50})_{\text{эксп}} = -0.03(\pm 0.04) + 1.11(\pm 0.05) \log(1/LD_{50})_{\text{расч}}$ $n = 234, r^2 = 0.714, s = 0.40, q^2 = 0.707, F = 579$
D	$\log(1/LD_{50})_{\text{эксп}} = 0.02(\pm 0.02) + 0.97(\pm 0.02) \log(1/LD_{50})_{\text{расч}}$ $n = 1522, r^2 = 0.596, s = 0.43, q^2 = 0.594, F = 2241$	$\log(1/LD_{50})_{\text{эксп}} = 0.10(\pm 0.03) + 0.91(\pm 0.04) \log(1/LD_{50})_{\text{расч}}$ $n = 380, r^2 = 0.551, s = 0.48, q^2 = 0.544, F = 464$
E	$\log(1/LD_{50})_{\text{эксп}} = 0.07(\pm 0.01) + 0.87(\pm 0.02) \log(1/LD_{50})_{\text{расч}}$ $n = 2017, r^2 = 0.476, s = 0.46, q^2 = 0.474, F = 1830$	$\log(1/LD_{50})_{\text{эксп}} = 0.06(\pm 0.03) + 0.85(\pm 0.04) \log(1/LD_{50})_{\text{расч}}$ $n = 509, r^2 = 0.419, s = 0.52, q^2 = 0.413, F = 365$
F	$\log(1/LD_{50})_{\text{эксп}} = 0.11(\pm 0.02) + 0.95(\pm 0.02) \log(1/LD_{50})_{\text{расч}}$ $n = 2862, r^2 = 0.306, s = 0.65, q^2 = 0.304, F = 1258$	$\log(1/LD_{50})_{\text{эксп}} = 0.18(\pm 0.03) + 0.81(\pm 0.04) \log(1/LD_{50})_{\text{расч}}$ $n = 717, r^2 = 0.364, s = 0.60, q^2 = 0.358, F = 409$

Результаты применения каскадной модели AMP для обучающего и тестового ряда соединений (табл. 28) подтверждают стабильность моделей. Очевидно, что стандартные отклонения уравнений на каждом уровне приблизительно те же самые, что и у общего ряда данных. Кроме того, на уровне А-Е они являются удовлетворительными и сопоставимыми с ошибками экспериментальных измерений. Тем не менее, необходимо отметить, что точность предсказания на основе каскадной модели AMP зависит как от качества токсикологических данных, так и от числа и качества структурных соседей. Подход AMP работает лучше на больших наборах данных.

5.3. Модель ЛРМПК

Ранее уже отмечалось, что одним из путей при использовании QSAR является развитие специфических локальных моделей для кластеров структурно-родственных соединений. В литературе описано несколько подобных подходов с применением методологии дискретно-непрерывных моделей (например, SIMCA/PLS [389], кластерно-регрессионные модели [96, 390]), которые успешно применены в расчете различных свойств/активности различных соединений. Однако, во всех этих подходах изученные выборки разделялись на ограниченное число кластеров. И, в случае обширных выборок, такие кластеры содержат большое число разнообразных химических соединений. Очевидно, в таких случаях необходима разработка иных моделей QSAR.

В качестве одного из решений этой проблемы можно рассматривать локальную отложенную регрессию (local lazy regression). В этом подходе, описанном в работах [391, 392], на основе метода *k*-ближайших соседей (KNN) производится кластеризация, и в каждом кластере строится собственная регрессионная модель.

В настоящей работе в качестве дальнейшего развития концепции дискретно-непрерывных QSAR моделей предлагается подход на основе регрессионных моделей в перекрывающихся кластерах. Указанный подход применен к расчету острой внутривенной токсичности по отношению к мышам для 10241 органического соединения.

Алгоритм построения локальных регрессионных моделей в перекрывающихся кластерах (ЛРМПК) может быть представлен следующим образом:

Шаг 1. Для рассматриваемого (i -го) соединения из выборки, содержащей n соединений, рассчитываются меры структурного сходства (коэффициенты Танимото T_c) с остальными $n-1$ соединениями. Полученные величины ранжируются по убыванию. Процедура повторяется для каждого соединения. В результате получается n отдельных рядов соединений.

Шаг 2. Фиксируется минимальный порог структурного сходства T_c , и в каждом из n рядов отбираются структурно-родственные соединения, имеющие T_c выше фиксированного уровня. Таким образом, формируются n перекрывающихся кластеров (рис. 20).

Шаг 3. Для всех анализируемых соединений рассчитываются дескрипторы.

Шаг 4. Для каждого кластера производится отбор дескрипторов на основе парных коэффициентов линейной корреляции ($r \leq 0.4$) и генерируются отдельные локальные регрессионные QSAR модели с использованием полного перебора всех возможных комбинаций из одного ($M=1$), двух ($M=2$) или трех ($M=3$) дескрипторов. Выбирается лучшее уравнение на основе критерия Фишера (F). В работе использовались линейные регрессионные модели при наличии в кластере ($5M + 1$) и более структурных соседей. Это обеспечивает наличие минимум 5 соединений на один дескриптор в процессе перекрестного контроля с исключением по одному. Важно отметить, что рассматриваемое соединение никогда не включается в построение этих регрессионных уравнений.

Шаг 5. Полученные в шаге 4 QSAR уравнения используются для расчета активности рассматриваемых соединений.

Таким образом, в рамках данного подхода каждое рассматриваемое соединение является тестовым соединением. А качество финальных результатов использования описанных моделей определяется размером и качеством выборки, наличием достаточного количества близких структурных соседей в каждом кластере, стабильностью построенных локальных QSAR моделей. Фактически, представленный выше алгоритм модели ЛРМПК, является логическим развитием регрессионной модели с учетом свойств ближайшего соседа/соседей, описанной в публикации [393]. Различие этих двух моделей связано только со способом использования информации о ближайших структурных соседях. Если в предыдущей модели свойства ближайших соседей использовались по отдельности для расчета свойства рассматриваемого соединения на основе общих коэффициентов регрессионного уравнения, то в настоящем подходе ближайшие структурные соседи используются для построения собственной локальной регрессионной модели кластера.

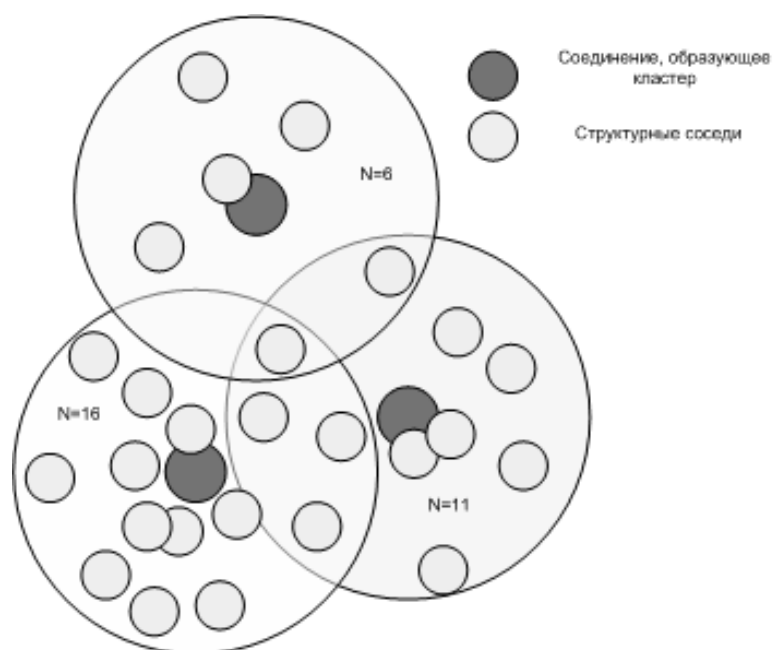


Рис. 20. Графическое представление метода ЛРМПК

Табл. 29 содержит результаты расчетов острой внутривенной токсичности 10241 соединения по отношению к мышам с помощью ЛРМПК подхода на основе НУВОТ дескрипторов.

Таблица 29

Параметры, статистические характеристики уравнения $\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ при различных уровнях структурного сходства (T_c) и распределение числа моделей при использовании одного ($M=1$), двух ($M=2$) и трех ($M=3$) дескрипторов

№	T_c	a_0	a_1	n	r^2	s	q^2	F	M=1	M=2	M=3
1	≥ 0.3	0.01(± 0.01)	0.99(± 0.01)	7759	0.434	0.51	0.434	5958	6425	1176	158
2	≥ 0.4	0.06(± 0.01)	0.91(± 0.01)	5011	0.530	0.46	0.529	5646	4074	810	127
3	≥ 0.5	0.06(± 0.01)	0.90(± 0.01)	3157	0.621	0.40	0.620	5170	2514	540	103
4	≥ 0.6	0.09(± 0.01)	0.87(± 0.02)	1748	0.621	0.40	0.618	2862	1457	260	31
5	≥ 0.7	0.15(± 0.02)	0.76(± 0.02)	924	0.618	0.38	0.613	1490	814	99	11
6	≥ 0.8	0.10(± 0.03)	0.88(± 0.03)	346	0.659	0.37	0.643	666	272	74	0

Из данных табл. 29 (№ 1), в частности, следует, что при задании в качестве граничного значения $T_c \geq 0.3$ и с учетом необходимого минимального числа структурных соседей, было получено 7759 кластеров, для которых было сконструировано 6425 однопараметровых, 1176 двухпараметровых и 158 трехпараметровых регрессионных уравнений для расчета токсичности соответствующих 7759 соединений. Для оставшихся 2482 соединений не было обнаружено достаточного числа структурных соседей, необходимого для оценки токсичности с помощью описанного метода. Статистические характеристики корреляции между экспериментальными и рассчитанными на основе соответствующих локальных регрессионных уравнений значениями токсичности для 7759 соединений оказываются очень хорошими. В указанном финальном уравнении свободный член и угловой коэффициент оказались идеальными (0.01 ± 0.01 и 0.99 ± 0.01 соответственно), а стандартное отклонение - на уровне ошибки экспериментального измерения этого свойства для мышей разного возраста, пола, условий их содержания и условий измерений.

Использование порога структурного сходства на уровне $T_c \geq 0.4$ (табл. 29, № 2) существенно уменьшило размер выборки (с 7759 до 5011 соединений) и

количество структурных соседей в кластере. Параметры и статистические критерии корреляции между экспериментальными и рассчитанными значениями $\log(1/LD_{50})$ стали несколько лучше. Эта тенденция была обнаружена и в процессе дальнейшего увеличения порога структурного сходства: при $T_c \geq 0.50$ – 3157 уравнений; при $T_c \geq 0.60$ – 1748 уравнений; при $T_c \geq 0.70$ – 924 уравнений; при $T_c \geq 0.80$ – 346 уравнений.

Мы провели анализ распределения дескрипторов среди регрессионных уравнений с одним, двумя и тремя независимыми переменными (рис. 21). При этом три дескриптора HUBOT: $\sum C_d/\alpha$ (№29) - композитный дескриптор, характеризующий свободноэнергетический протонодонорный дескриптор на единицу молекулярной поляризуемости; $\sum E_d/\alpha$ (№26) - композитный дескриптор, характеризующий энтальпийный протонодонорный дескриптор на единицу молекулярной поляризуемости и молекулярная поляризуемость α (№1) оказались наиболее используемыми в однопараметровых уравнениях (20%, 18% и 9% всех однопараметровых уравнений соответственно).

Парная комбинация данных дескрипторов и свободноэнергетических или энтальпийных протоноакцепторных дескрипторов водородной связи ($\sum C_a$, $\sum E_a$) довольно распространена в двухпараметровых уравнениях. Необходимо заметить, что эти дескрипторы HUBOT напрямую связаны с транспортными свойствами (растворимость в воде, липофильность и т.д.) химических соединений и лекарств. Таким образом, можно предположить, что транспортные свойства химических соединений и лекарств существенно влияют на их острую внутривенную токсичность.

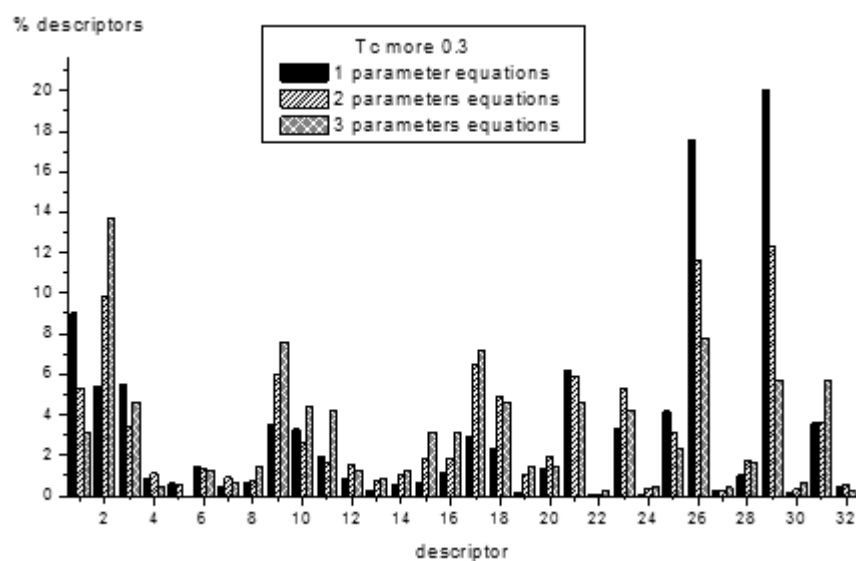


Рис. 21. Распределение НУВОТ дескрипторов в одно-, двух- и трехпараметровых уравнениях при $T_c \geq 0.3$ и взаимной корреляции дескрипторов на уровне $r \leq 0.4$ (1 – α , 2 – $\max(Q^+)$, 3 – $\max(Q^-)$, 4 – ΣQ^+ , 5 – ΣQ^- , 6 – $\Sigma |Q|$, 7 – $\Sigma Q^+/\alpha$, 8 – $\Sigma Q^-/\alpha$, 9 – $\max(E_a)$, 10 – $\max(C_a)$, 11 – $\max(C_{a(o)})$, 12 – $\max(E_d)$, 13 – $\max(C_d)$, 14 – $\max(E_a) \cdot \max(E_d)$, 15 – $\max(C_a) \cdot \max(C_d)$, 16 – $\max(C_{a(o)}) \cdot \max(C_d)$, 17 – ΣE_a , 18 – ΣE_d , 19 – ΣE_{ad} , 20 – ΣC_a , 21 – ΣC_d , 22 – ΣC_{ad} , 23 – $\Sigma C_{a(o)}$, 24 – $\Sigma C_{ad(o)}$, 25 – $\Sigma E_a/\alpha$, 26 – $\Sigma E_d/\alpha$, 27 – $\Sigma E_{ad}/\alpha$, 28 – $\Sigma C_a/\alpha$, 29 – $\Sigma C_d/\alpha$, 30 – $\Sigma C_{ad}/\alpha$, 31 – $\Sigma C_{a(o)}/\alpha$, 32 – $\Sigma C_{ad(o)}/\alpha$)

Также были проведены расчеты токсичности 10241 органического соединения с помощью подхода ЛРМПК на основе дескрипторов, рассчитанных компьютерной программой DRAGON. Зависимость значений коэффициента корреляции и стандартного отклонения от величины порога структурного сходства в уравнениях сравнения экспериментальных и рассчитанных значений $\log(1/LD_{50})$ представлены на рис. 22.

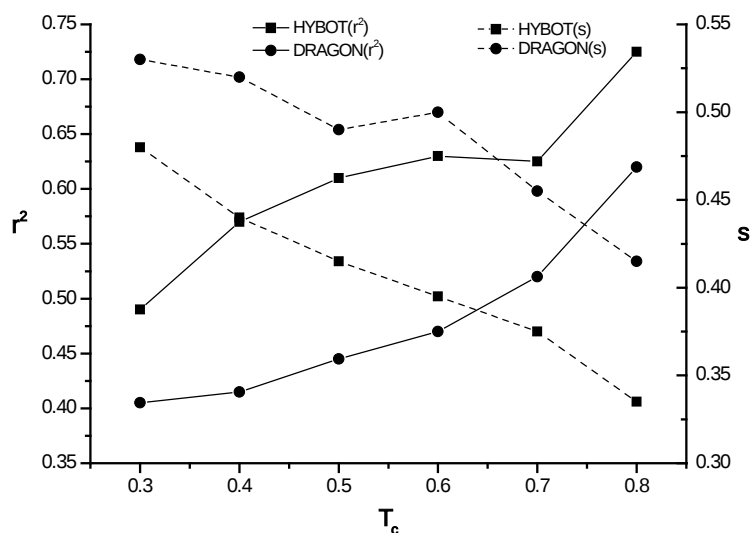


Рис. 22. Зависимость значений коэффициента корреляции (r^2) и стандартного отклонения (s) моделей от величины порога структурного сходства (T_c) при использовании HYBOT и DRAGON дескрипторов

Очевидно, что порог структурного сходства имеет существенное влияние на точность расчетов. Статистические критерии таких уравнений становятся лучше при большем значении данного порога. Но одновременно количество получаемых уравнений, а, следовательно, и число соединений с рассчитанным значением токсичности уменьшается. Это явно демонстрирует эффективность использования концепции структурного сходства для оценки свойств/активности соединений. Рис. 22 также демонстрирует, что лучшие результаты расчетов получаются при использовании HYBOT дескрипторов.

Для валидации модели были сформированы обучающая (8193 соединения) и тестовая (2048 соединений) выборки. Результаты применения ЛРМПК модели с использованием HYBOT дескрипторов представлены в табл. 30. Из представленных данных очевидна стабильность полученных ЛРМПК моделей и эффективность использования концепции структурного сходства для расчета токсичности соединений.

Таблица 30

Параметры и статистические характеристики уравнения $\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ для обучающей (О) и тестовой (Т) выборок при различных уровнях структурного сходства (T_c)

T_c	Ряд	a_0	a_1	r^2	s	q^2	F	n
≥ 0.3	О	0.01(± 0.01)	0.98(± 0.01)	0.428	0.52	0.428	4403	5876
	Т	0.05(± 0.02)	0.92(± 0.03)	0.398	0.53	0.394	972	1469
≥ 0.4	О	0.04(± 0.01)	0.93(± 0.01)	0.553	0.45	0.552	4455	3604
	Т	0.07(± 0.02)	0.89(± 0.03)	0.538	0.46	0.535	1056	908
≥ 0.5	О	0.09(± 0.02)	0.84(± 0.02)	0.576	0.42	0.574	3011	2222
	Т	0.09(± 0.02)	0.83(± 0.03)	0.590	0.43	0.577	803	560

5.4. Проницаемость ФАВ через гематоэнцефалический барьер

Общеизвестна та важная роль, которую играет гематоэнцефалический барьер (ГЭБ) для обеспечения нормальной работы центральной нервной системы [424]. Несмотря на большое количество работ, посвященных исследованию ГЭБ, связь между структурой низкомолекулярных биорегуляторов и их способностью проникать через барьер остается недостаточно выявленной.

Важным механизмом, обеспечивающим транспорт ФАВ через ГЭБ, является пассивная диффузия. Для количественного моделирования этого механизма нами была использована модель АМР, описанная ранее. В основу данных были положены публикации [425, 426], содержащие сведения о $\log BV$ (десятичный логарифм отношения концентраций вещества в мозге и крови при пассивной диффузии через ГЭБ) для 298 соединений. С целью валидации модели исходные 298 соединений разбивались на обучающую (239 соединений) и тестовую (59 соединений) выборки. В качестве внешней тестовой выборки использовались также данные по $\log BV$ для пяти дополнительных соединений, не включенных в общую выборку [427]. Для описания структуры ФАВ использовали 32 дескриптора, полученных на основе программы HUYBOT [305].

В качестве меры структурного сходства фигурировали коэффициенты T_c , рассчитанные с помощью программы MOLDIVS [323].

Схема расчета величин $\log BB$ представлена на рис. 23. Для создания модели АМР необходимо рассчитать матрицы дескрипторов и структурного сходства, сформировать обучающую и тестовую выборки, ввести необходимые управляющие параметры. На основе обучающей выборки путем перебора комбинаций дескрипторов генерируется ряд моделей. С использованием статистических критериев выбирается лучшая модель и на ее основе рассчитываются величины $\log BB$ для молекул тестового ряда.

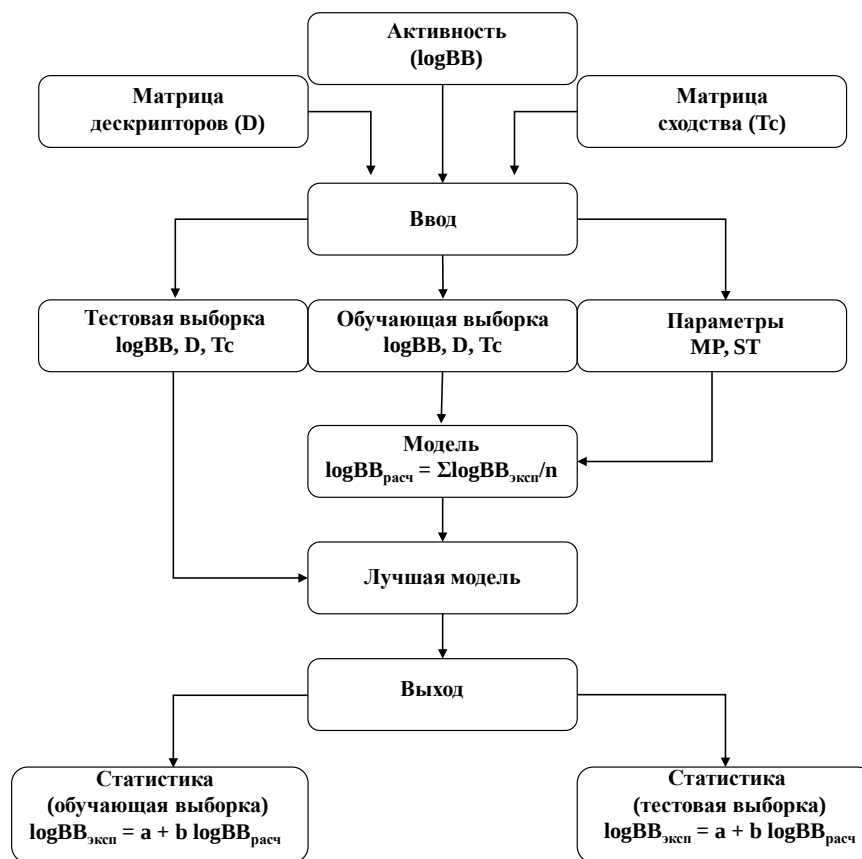


Рис. 23 Принципиальная схема расчета величин проницаемости ($\log BB$) ФАВ через ГЭБ (D-дескрипторы, T_c -коэффициент Танимото, MP-число молекулярных пар, ST-порог структурного сходства)

В результате проведенной работы были выделены 3 лучшие модели на основе одиночных дескрипторов ΣC_{ad} – сумма свободноэнергетических Н-акцепторных и Н-донорных дескрипторов, композитный дескриптор $\Sigma C_{ad}/\alpha$, где α - молекулярная поляризуемость и ΣE_{ad} – сумма энтальпийных Н-акцепторных и Н-донорных дескрипторов. При использовании в качестве обучающей выборки массива из 298 соединений для вышеперечисленных моделей были получены величины r^2 и s равные 0.594 и 0.41, 0.588 и 0.41, 0.577 и 0.42 соответственно.

Как указывалось выше, в указанных моделях используются одиночные дескрипторы. На финальной стадии работы можно получить согласованную

модель с использованием в качестве конечного результата среднего арифметического рассчитанных значений из 3-х моделей по отдельным дескрипторам. Применение такого подхода приводит к получению модели со следующими статистическими характеристиками: $r^2=0.641$ и $s=0.38$. Последняя цифра близка к значению s равному 0.37, полученному при корреляции экспериментальных значений $\log BB$ одних тех же соединений, взятых из разных источников. Эта величина хорошо согласуется с литературными данными, согласно которым граница варибельности величин $\log BB$, полученных из разных источников, составляет 0.5. ~~Логарифмических~~ единиц [427]. Это свидетельствует о том, что полученная нами финальная модель имеет стандартное отклонение на уровне ошибки экспериментального определения этого параметра.

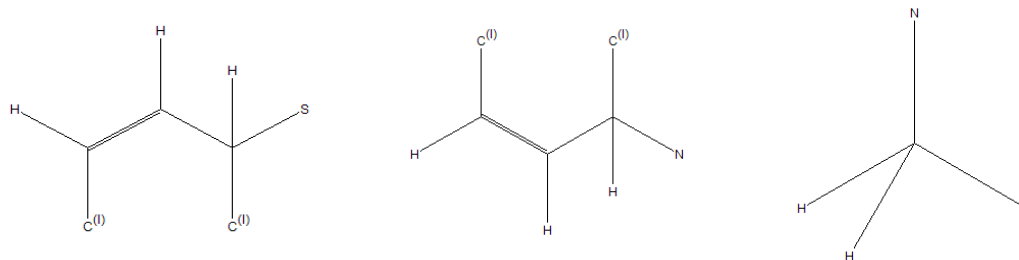
На заключительном этапе для тестирования описанной схемы расчета были привлечены данные по проницаемости пяти дополнительных соединений. Это три вещества с пассивным механизмом транспорта (лоперамид, хлорпромазин, клонидин) и два вещества с активным механизмом транспорта (терфенадин, ранитидин). Результаты расчета приведены в табл. 31. Из представленных данных следует, что не только соединения с пассивным механизмом транспорта, но и соединения, имеющие существенный вклад активного транспорта, хорошо описываются предложенной моделью. Это позволяет предположить, что разработанный метод расчета может быть применен к химическим соединениям, проникающим через ГЭБ с помощью различных механизмов.

Таблица 31

Экспериментальные ($\log BB_{\text{эксп}}$) и рассчитанные ($\log BB_{\text{расч}}$) величины проницаемости ФАВ через ГЭБ на основе различных АМР моделей с использованием дескрипторов: $\sum E_{\text{ad}}$ – сумма энтальпийных Н-акцепторных и Н-донорных дескрипторов, $\sum C_{\text{ad}}$ – сумма свободноэнергетических Н-акцепторных и Н-донорных дескрипторов, α – молекулярная поляризуемость ($\text{\AA}^3$).

№	ФАВ	$\log BB_{\text{эксп}}$	$\log BB_{\text{расч}}$			
			$\sum E_{\text{ad}}$	$\sum C_{\text{ad}}$	$\sum C_{\text{ad}}/\alpha$	$\sum E_{\text{ad}} + \sum C_{\text{ad}} + \sum C_{\text{ad}}/\alpha$
1	Лоперамид	0.77	0.45	0.45	0.94	0.61
2	Хлорпромазин	0.84	0.70	1.02	1.02	0.91
3	Клонидин	0.19	-0.14	0.25	0.25	0.12
4	Терфенадин	0.64	0.45	0.45	0.70	0.53
5	Ранитидин	-1.23	-1.27	-1.03	-1.03	-1.11

На основе глобальных моделей «структура – активность», представленных в главе 5, может проводиться компьютерный дизайн органических соединений путем решения прямой и обратной задачи. Пример успешного решения прямой задачи на основе модели АМР представлен в табл. 31. Обратная задача (предсказание структур с заданными свойствами) из-за своей некорректности может быть решена путем введения дополнительных ограничений. Проанализируем, например, соединения с высокой величиной проницаемости через ГЭБ. Введем следующие ограничения: $\log BB$: $1.0 \div 1.5$; $T_c \geq 0.5$; $NP=1$; дескриптор: $\sum C_{\text{ad}}$. В результате на основе обучающей выборки был установлен набор из 30 молекулярных фрагментов, 19 из которых являются общими. Некоторые из этих фрагментов изображены ниже:



При изменении ограничений модели AMP набор общих молекулярных фрагментов также меняется.

Выводы по главе

С использованием параболических и билинейных уравнений а также величин липофильности, молекулярной поляризуемости и дескрипторов Н-связи, получены обладающие высокими статистическими характеристиками модели базовой токсичности 46 органических соединений по отношению к мышам.

Выявлено, что при проведении бинарной классификации 9618 органических соединений, обладающих острой токсичностью по отношению к мышам, разработанный в настоящей работе метод АЛБТ, основанный на использовании двух дескрипторов: молекулярной поляризуемости и Н-акцепторной способности, показывает более высокие характеристики по сравнению с методами LDA, KNN, SVM, ANN, SS и обеспечивает степень правильного распознавания в обучающей и тестовой выборке на уровне 0.998 и 0.824 соответственно.

Продемонстрирована применимость AMP подхода для предсказания острой токсичности 10241 соединения по отношению к мышам. Установлено, что использование каскадной модели AMP с дескриптором $\sum C_{ad}/\alpha$ позволило оценить острую токсичность 7085 соединений (около 69% от анализируемого ряда данных) со стандартным отклонением, сравнимым с экспериментальной ошибкой определения.

На основе концепции структурного сходства и локальных регрессионных уравнений, разработана ЛРМПК модель, позволяющая рассчитывать токсичность больших массивов органических соединений разнообразных химических классов. Указанный подход успешно применен к расчету токсичности 7759 соединений, каждое из которых имеет в рассматриваемой выборке достаточное количество структурно-родственных соединений для построения собственной локальной регрессионной модели соответствующего кластера. Два дескриптора НУВОТ: $\sum C_d/\alpha$ и $\sum E_d/\alpha$ оказались наиболее используемыми в однопараметровых уравнениях.

Выявлено, что модель АМР может быть с успехом использована для расчета величины пассивной проницаемости ФАВ через ГЭБ. Установлено, что величины стандартных отклонений уравнений соответствуют величинам вариабельности $\log BB$, полученным из разных источников. Показано, что модель АМР может быть использована для оценки величин проницаемости ФАВ в случае различных механизмов действия.

Глава 6. Перспективные разработки в области создания количественных моделей «структура – свойство» и «структура – активность»

6.1. Рекуррентная модель

Гомологические ряды органических соединений находят применение, в частности, в качестве объектов при изучении механизмов токсичности [2, 382]. Это связано с тем, что гомологи являются структурно близкими соединениями, имеющими одинаковый тип действия. Известно, что в гомологических рядах молекул с ростом числа атомов углерода токсичность увеличивается до известного предела, после которого наступает насыщение или даже разворот [169, 384]. Это может быть связано с различными процессами, например, с уменьшением растворимости веществ или ростом степени ассоциации молекул.

Для описания зависимости токсичности химических соединений от их структуры используются как простые линейные, так и параболические модели [2, 7, 184, 185]. Очевидно, что линейные уравнения имеют ограниченную область применения для гомологических рядов молекул. Параболические модели выглядят более предпочтительными, но их надежность при экстраполяции данных вызывает сомнения. Не так давно была продемонстрирована эффективность использования рекуррентной модели для описания различных физико-химических свойств молекул в гомологических рядах [385].

В настоящей работе на основе рекуррентных уравнений проведено исследование острой токсичности 85 органических соединений из 7 гомологических рядов по отношению к бактерии *Vibrio fisheri*, инфузории *Tetrahymena pyriformis*, рыбам *Poecilia reticulata* и *Pimephales promelas*, мышам и крысам (внутривенно).

В табл. 32 представлены данные по острой токсичности исследованных соединений, включая спирты (13 соединений), кетоны (11), сложные эфиры

(10), амины (12), фенолы (10), альдегиды (12) и кислоты (17). Токсичность меняется в широких пределах: для водных организмов от $\log(1/LC_{50})=-2.96$ до $\log(1/EC_{50})=3.69$. Для теплокровных организмов минимальное значение $\log(1/LD_{50})$ составляет -2.17, а максимальное 1.09.

Зависимость токсичности от числа атомов углерода во всех исследованных гомологических рядах имеет прямолинейный характер или стремится к некоторому пределу. Поэтому для аппроксимации данных может быть использовано простое рекуррентное уравнение вида:

$$T_n = A + BT_{n-1} \quad (152)$$

где T_n – токсичность текущего n -го члена гомологического ряда, T_{n-1} – токсичность предшествующего члена гомологического ряда, A и B – параметры.

Результаты соответствующих расчетов представлены в табл. 33. Полученные уравнения характеризуются высокими статистическими показателями. Например, в большинстве случаев, $r^2 > 0.9$. Часть представленных зависимостей имеет предельное значение токсичности ($B < 1$), другая часть не имеет предела ($B > 1$). При этом коэффициент A варьирует от 0.13 до 0.67, а коэффициент B варьирует от 0.81 до 1.20.

Таблица 32

Острая токсичность органических соединений по отношению к различным биологическим объектам

№	Соединение	Биологический объект					
		V.f.	T.p.	P.r.	P.p.	M	K
1	Метанол	-	-2.67	-2.94	-2.96	-2.17	-1.82
2	Этанол	-2.89	-1.99	-2.44	-2.50	-1.63	-1.49
3	1-Пропанол	-2.22	-1.75	-1.95*	-1.88	-1.06	-0.99
4	1-Бутанол	-1.66	-1.43	-1.37	-1.37	-0.71	-0.62
5	1-Пентанол	-0.53	-1.05	-0.67*	-0.73	-0.32	-0.28
6	1-Гексанол	0.68	-0.38	0.02	0.02	0.00	-
7	1-Гептанол	1.07	0.11	0.55*	0.53	0.19*	-
8	1-Октанол	1.58	0.63	0.98	0.98	0.28	-
9	1-Нонанол	1.87*	0.85	1.40	1.40	-	-
10	1-Деканол	2.08	1.34	1.81	1.82	-	-
11	1-Ундеканол	2.59*	1.95	2.21	-	-	-
12	1-Додеканол	3.69	2.16	2.26	-	-	-
13	1-Тридеканол	-	2.45		-	-	-
14	Ацетон	-2.44	-2.15	-2.10	-2.09	-	-
15	2-Бутанон	-1.68	-1.75	-1.65	-1.65	-	-
16	2-Пентанон	-0.93*	-1.22	-1.13*	-1.16	-	-
17	2-Гексанон	-0.23*	-1.34	-0.56*	-0.63	-	-
18	2-Гептанон	0.38*	-0.49	0.01*	-0.06	-	-
19	2-Октанон	0.86	-0.15	0.55	0.55	-	-
20	2-Нонанон	1.18*	0.66	1.04*	0.97	-	-
21	2-Деканон	1.30	0.58	1.43	1.51	-	-
22	2-Ундеканон	-	1.53	-	2.06	-	-
23	2-Додеканон	-	1.67	-	2.19	-	-

Продолжение табл. 32

№	Соединение	Биологический объект					
		V.f.	T.p.	P.r.	P.p.	M	K
24	2-Тридеканон	-	2.12	-	2.74	-	-
25	Метилацетат	-	-1.60	-	-0.68	-	-
26	Этилацетат	-1.77	-1.30	-	-0.42	-	-
27	Пропилацетат	-	-1.24	-	0.23	-	-
28	Бутилацетат	0.22	-0.49	-	0.81	-	-
29	Амилацетат	-	-0.15*	-	1.25*	-	-
30	Гексилацетат	-	-0.01	-	1.52	-	-
31	Гептилацетат	-	0.44*	-	-	-	-
32	Октилацетат	-	1.06	-	-	-	-
33	Нонилацетат	-	1.59*	-	-	-	-
34	Децилацетат	-	1.88	-	-	-	-
35	Этиламин	-	-	-0.70	-	-	-
36	Пропиламин	-	-0.71	-0.72	-0.72	-	-
37	Бутиламин	-	-0.57	-0.56	-0.56	-0.43	-
38	Амиламин	-	-0.48	-0.31	-0.31	-0.11*	-
39	Гексиламин	-	-0.22	0.25	0.25	0.22*	-
40	Гептиламин	-	0.21	0.72	0.72	0.54*	-
41	Октиламин	-	0.35	1.40	1.40	0.86	-
42	Нониламин	-	1.70	1.82	1.82	-	-
43	Дециламин	-	2.06	2.18	2.18	-	-
44	Ундециламин	-	2.33	2.91	2.91	-	-
45	Додециламин	-	-	3.27	3.26	-	-
46	Тридециламин	-	-	3.46	3.48	-	-
47	Фенол	-	-0.35	0.45	0.46	-	-
48	4-Метилфенол	-	-0.16	0.74	0.82	-	-
49	4-Этилфенол	-	0.21	1.07	1.07	-	-

Продолжение табл. 32

№	Соединение	Биологический объект					
		V.f.	T.p.	P.r.	P.p.	M	K
50	4-Пропилфенол	-	0.64	1.09	1.09	-	-
51	4-Бутилфенол	-	0.88*	1.47	1.62*	-	-
52	4-Пентилфенол	-	1.20*	2.12	1.92*	-	-
53	4-Гексилфенол	-	1.52*	2.24*	2.22*	-	-
54	4-Гептилфенол	-	1.84*	2.55*	2.52*	-	-
55	4-Октилфенол	-	2.16*	2.86*	2.82*	-	-
56	4-Нонилфенол	-	2.47	3.20	3.20	-	-
57	Метаналь	-	-	0.04	-	-	-
58	Этаналь	-	-	0.10	0.12	-	-
59	Пропаналь	-	-0.49	0.59	-	-	-
60	Бутаналь	-	-0.38	0.72	0.69	-	-
61	Пентаналь	-	-0.02	0.82	0.82	-	-
62	Гексаналь	-	-0.17	1.01	0.76	-	-
63	Гептаналь	-	0.00	1.11	-	-	-
64	Октаналь	-	0.48	1.21	-	-	-
65	Нонаналь	-	0.81	1.38*	-	-	-
66	Деканаль	-	1.28	1.69	-	-	-
67	Ундеканаль	-	1.69	-	-	-	-
68	Додеканаль	-	1.76	-	-	-	-
69	Уксусная кислота	-	-	-	-	-0.94	-
70	Пропановая кислота	-	-0.51	-	-	-0.93	-
71	Бутановая кислота	-	-0.57	-	-	-0.96	-
72	Пентановая кислота	-	-0.27	-	-	-1.10	-
73	Гексановая кислота	-	-0.21	-	-	-1.17	-
74	Гептановая кислота	-	-0.11	-	-	-0.96	-
75	Октановая кислота	-	0.08	-	-	-0.62	-

Продолжение табл. 32

№	Соединение	Биологический объект					
		V.f.	T.p.	P.r.	P.p.	M	K
76	Нонановая кислота	-	0.35	-	-	-0.15	-
77	Декановая кислота	-	0.51	-	-	0.13	-
78	Ундекановая кислота	-	0.90	-	-	0.12	-
79	Додекановая кислота	-	-	-	-	0.18	-
80	Тридекановая кислота	-	-	-	-	0.22	-
81	Тетрадекановая кислота	-	-	-	-	0.73	-
82	Пентадекановая кислота	-	-	-	-	0.65	-
83	Гексадекановая кислота	-	-	-	-	0.65	-
84	Гептадекановая кислота	-	-	-	-	0.88	-
85	Октадекановая кислота	-	-	-	-	1.09	-

Примечание. V.f. - *Vibrio fisheri* ($\log(1/EC_{50})$, ммоль/л); T.p. - *Tetrahymena pyriformis* ($\log(1/IGC_{50})$, ммоль/л); P.r. - *Poecilia reticulata* ($\log(1/LC_{50})$, ммоль/л); P.p. - *Pimephales promelas* ($\log(1/LC_{50})$, ммоль/л); M - мыши ($\log(1/LD_{50})$, ммоль/кг); K - крысы ($\log(1/LD_{50})$, ммоль/кг); * - интерполированные значения на основе кубических сплайнов или регрессионных уравнений.

Таблица 33

Параметры и статистические характеристики рекуррентных уравнений

№	Соединения	Биообъект	A	B	n	r ²	s	F	q ²
1	Спирты	V. fisheri	0.67(±0.12)	0.95(±0.06)	10	0.966	0.37	226.1	0.946
2	Спирты	T. pyriformis	0.42(±0.05)	0.97(±0.03)	12	0.989	0.17	879.0	0.983
3	Спирты	P. reticulata	0.46(±0.04)	0.94(±0.03)	11	0.994	0.14	1395.5	0.989
4	Спирты	P. promelas	0.51(±0.04)	0.97(±0.03)	9	0.995	0.11	1470.5	0.992
5	Спирты	Мыши	0.20(±0.04)	0.81(±0.03)	7	0.992	0.07	634.3	0.982
6	Спирты	Крысы	0.34(±0.13)	0.97(±0.10)	4	0.978	0.09	90.0	0.907
7	Кетоны	V. fisheri	0.47(±0.04)	0.84(±0.03)	7	0.992	0.11	645.8	0.981
8	Кетоны	T. pyriformis	0.42(±0.13)	0.97(±0.10)	10	0.926	0.39	99.7	0.896
9	Кетоны	P. reticulata	0.49(±0.03)	0.98(±0.02)	7	0.997	0.07	1633.4	0.992
10	Кетоны	P. promelas	0.49(±0.04)	0.98(±0.03)	10	0.992	0.14	1026.4	0.986
11	Сл. эфиры	T. pyriformis	0.39(±0.08)	1.03(±0.08)	9	0.963	0.24	181.9	0.939
12	Сл. эфиры	P. promelas	0.45(±0.09)	0.94(±0.12)	5	0.952	0.20	60.1	0.842
13	Амины	T. pyriformis	0.36(±0.16)	1.05(±0.16)	8	0.882	0.44	44.8	0.833
14	Амины	P. reticulata	0.34(±0.08)	1.04(±0.05)	11	0.980	0.23	431.1	0.968
15	Амины	P. promelas	0.41(±0.08)	1.01(±0.05)	10	0.983	0.20	459.7	0.972
16	Амины	Мыши	0.32(0.01)	1.00(0.01)	4	0.999	0.01	14951.0	0.999
17	Фенолы	T. pyriformis	0.31(±0.04)	1.01(±0.03)	9	0.994	0.07	1172.8	0.989
18	Фенолы	P. reticulata	0.30(±0.14)	1.00(±0.08)	9	0.960	0.19	168.4	0.946
19	Фенолы	P. promelas	0.27(±0.11)	1.02(±0.06)	9	0.975	0.14	270.0	0.962
20	Альдегиды	T. pyriformis	0.24(±0.08)	1.04(±0.10)	9	0.933	0.23	97.7	0.886
21	Альдегиды	P. reticulata	0.22(±0.10)	0.96(±0.11)	9	0.917	0.14	77.1	0.808
22	Кислоты	T. pyriformis	0.19(±0.05)	1.20(±0.13)	8	0.935	0.13	85.9	0.857
23	Кислоты	Мыши	0.13(±0.05)	1.02(±0.07)	16	0.934	0.21	199.6	0.917

Полезность математических моделей для описания токсичности, со статистической точки зрения, можно связать с качеством экстраполяции данных. С этой целью мы сопоставили три модели: рекуррентную (уравнение (152)), линейную:

$$T = a_0 + a_1 M_{CH_2} \quad (153)$$

и параболическую:

$$T = a_0 + a_1 M_{CH_2} + a_2 M_{CH_2}^2 \quad (154)$$

где T – токсичность; M_{CH_2} – число метиленовых групп в гомологическом ряду $R_1 - (CH_2)_m - R_2$; a_0, a_1, a_2 – параметры уравнений.

Для сравнения были выбраны данные по токсичности с учетом количества молекул в ряду и с максимальным представлением всех использованных типов химических соединений и биологических объектов. В результате было отобрано 5 комбинаций гомологический ряд/биообъект: спирты и *T. pyriformis*, кетоны и *V. fisheri*, амины и *P. promelas*, альдегиды и *P. reticulata*, кислоты и мыши. Из каждого ряда отбрасывались последние 3 точки, а на основе оставшихся соединений рассчитывались соответствующие уравнения. Затем эти уравнения использовались для расчета токсичности отброшенных точек. В качестве статистики для сравнения результатов было использовано стандартное отклонение s для экспериментальных и рассчитанных значений токсичности. Всего для каждой модели было получено по 5 значений s на основе 15 тестовых точек. В результате было установлено, что для линейной модели среднее значение $s=0.36$, для параболической модели $s=0.49$ и для рекуррентной модели $s=0.33$. Полученные результаты свидетельствуют о том, что рекуррентная модель обладает наилучшей способностью к экстраполяции среди рассмотренных трех уравнений.

Одним из эффективных подходов при исследовании связи «структура – токсичность» является использование классификации молекул по типу действия [2]. В литературе можно найти различные схемы классификации, при этом соединения одного типа могут называться по-разному, например,

«неполярный наркоз» [16] и «наркоз I» [238]. Обобщая [276], все типы действия можно разбить на три класса. Первый класс - неполярная неспецифическая токсичность, второй класс - полярная неспецифическая токсичность и третий класс - специфическая токсичность. Учитывая это, рассматриваемые гомологические ряды спиртов, кетонов и сложных эфиров можно отнести к классу 1, амины и фенолы к классу 2, альдегиды и кислоты - к классу 3.

Использованная классификация основана на изучении токсичности химических соединений по отношению к водным организмам (экотоксичность). Однако, с известной долей вероятности, она может быть использована и для теплокровных мышей и крыс. Основанием для такого вывода является факт существования корреляции токсичности между рассмотренными выше биологическими объектами [7].

Представленные в табл. 33 рекуррентные модели могут быть выражены в виде точек на плоскости А-В с учетом типа гомологического ряда и вида биологического объекта. На рис. 24 изображены коэффициенты рекуррентных уравнений для различных типов химических соединений.

Этот рисунок содержит 23 точки, которые соответствуют уравнениям из табл. 33. На нем отчетливо видны три кластера. В первый кластер входят спирты, кетоны и сложные эфиры. Во второй кластер – амины и фенолы. Третий кластер состоит из альдегидов и кислот. Таким образом, представленный метод графического анализа позволяет различать гомологические ряды между собой, относя их к веществам с различным типом токсического действия.

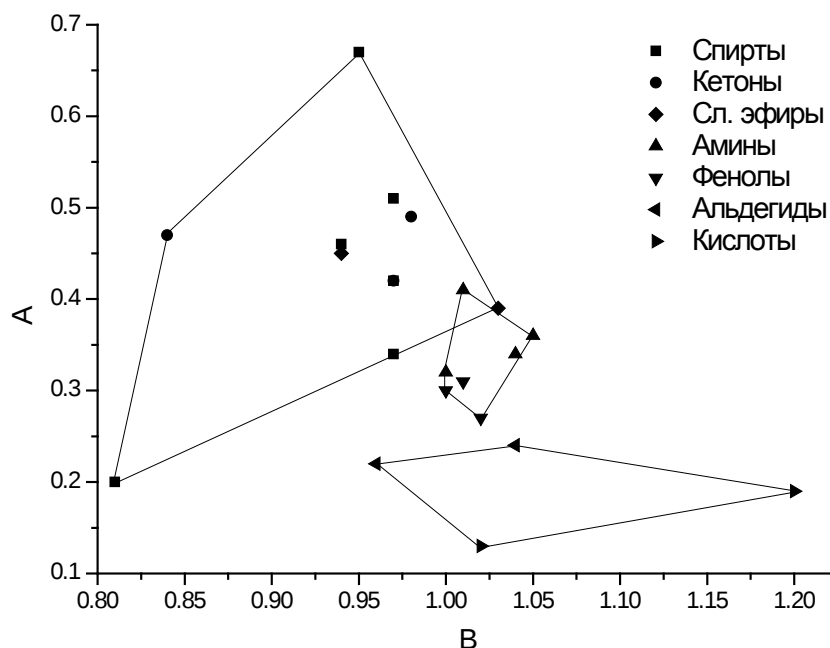


Рис. 24. Параметры рекуррентных уравнений на плоскости А-В с учетом типа гомологического ряда

Одной из важных дискутируемых тем в литературе является вопрос о сравнительной токсической чувствительности различных биологических объектов [7]. Одним из подходов для получения результатов является корреляционный анализ токсичности в одинаковых рядах соединений [198, 223]. Обычно, в этом случае, сходство биообъектов между собой оценивается с помощью коэффициента корреляции. В работе [386] использован другой подход, основанный на проекции данных из матрицы, образованной на основе токсичности одного и того же ряда соединений по отношению к различным биообъектам, на главные компоненты. Мы в своей работе анализировали токсичность различных гомологических рядов между различными биологическими объектами. При этом в качестве меры токсикологической близости биообъектов между собой использовали расстояние между точками на рис. 25.

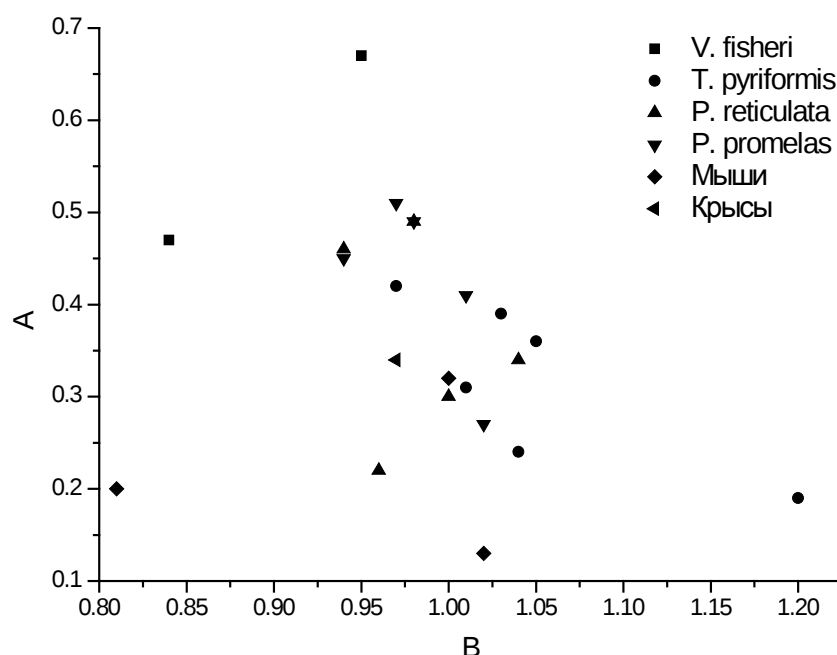


Рис. 25. Параметры рекуррентных уравнений на плоскости А-В с учетом вида биологического объекта

Обращает на себя внимание уникальность данных для бактерии *V. fisheri* ввиду отсутствия в ближайших окрестностях аналогичных моделей токсичности. Близость (№3, №12 из табл. 33) или даже совпадение (№ 9, №10) параметров рекуррентных уравнений, описывающих токсичность химических соединений по отношению к рыбам *P. reticulata* и *P. promelas* не вызывает удивления: все-таки это очень сходные организмы. В работе [22] установлена токсикологическая близость *T. pyriiformis* и *P. reticulata* по отношению к ряду химических соединений. Аналогично констатируется большое сходство регрессионных моделей, описывающих неспецифическую токсичность, по отношению к *T. pyriiformis* и *P. promelas* [62]. Из наших данных также следует сходство моделей токсичности *T. pyriiformis* и *P. reticulata* (№ 13, 14), а также *T. pyriiformis* и *P. promelas* (№ 11, 15).

В случае мышей две модели не имеют ближайших соседей. А вот параметры третьей модели, описывающей токсичность в ряду аминов (№ 16), оказываются очень близкими к моделям № 17 и № 18. Первая из них связана с токсичностью фенолов по отношению к *T. pyriformis*, а вторая описывает токсические свойства фенолов по отношению к *P. reticulata*. Что касается крыс (№ 6), то ближайшим биологическим соседом для них оказываются мыши (№ 16).

6.2. Фрактальный дескриптор

В последние десятилетия при исследовании различных объектов и процессов широкое распространение получила фрактальная геометрия [394, 395]. Концепция фракталов и связанное с ней понятие о фрактальной размерности в настоящее время успешно используются в различных областях естествознания [396-400]. В частности, при изучении свойств нанообъектов исследуются взаимоотношения между фрактальной размерностью и свойствами материалов [401, 402]. Существование таких связей, очевидно, обусловлено наличием фрактальных свойств у молекулярных кластеров наноразмера. При этом следует констатировать отсутствие аналогичных работ в области изучения взаимоотношений «структура – свойство» и «структура – активность», по крайней мере, при использовании обычных «малых» молекул. Вероятно, это связано с тем, что такие молекулы не обладают фрактальными свойствами. Это же отмечается в работе [403], которая является одной из немногих, в которой использованы фрактальные идеи для изучения связи «структура – активность». Следует отметить, что отсутствие инвариантности при изменении масштаба (самоподобия) геометрических структур молекул, выраженных, например, в виде декартовых координат атомов, не исключает присутствия самоподобия при других способах представления пространственной структуры. На основе этого может быть разработан новый 3D дескриптор молекулярной структуры – фрактальная размерность гистограммы межатомных расстояний.

Гистограммы (дискретные функции распределения) межатомных расстояний представлялись в виде плоских диаграмм, состоящих из набора закрашенных прямоугольников (баров). Ось абсцисс разбивалась с определенным шагом (разрешением) ΔR на интервалы, количество которых зависело от значений межатомных расстояний R , меняющихся от минимального значения $R_{\min}$ до максимального значения $R_{\max}$. На оси ординат откладывалась частота f , т.е. количество межатомных расстояний, попадающих в заданный интервал, и варьирующая в виде целых чисел от 0 до $f_{\max}$. Эта ось также разбивалась на интервалы с определенным шагом Δf . Полученные гистограммы оцифровывались путем их представления в виде квадратных бинарных матриц, где значение 1 соответствовало закрашенным участкам гистограммы, а 0 – всему остальному.

Фрактальная размерность определялась «клеточным» методом с помощью компьютерной программы, алгоритм которой приведен в работе [404], путем покрытия оцифрованного изображения клетками переменной величины и подсчетом числа клеток, которые пересекают закрашенные области гистограмм. Локальная фрактальная размерность рассчитывалась по уравнению:

$$\log N(L) = \log C - D \log L \quad (155)$$

где: $N(L)$ – минимальное число клеток со стороной L , необходимых для покрытия фрактала; D – «клеточная» фрактальная размерность (со знаком -); C – константа.

Для определения величины D , в соответствии с уравнением (155), нужно провести два измерения при двух различных величинах L . Однако, для более надежной оценки фрактальной размерности, необходимо проводить большее число измерений. С этой целью был использован набор клеток с размерами

сторон $L = 1, 2, \dots, L_{\max}$. Для каждой молекулы величина $L_{\max}$ определялась путем деления на 10 минимального значения из двух чисел: $(R_{\max} - R_{\min})/\Delta R$ и $f_{\max}/\Delta f$.

При расчете D необходимо учитывать то, что гистограммы имеют различные коэффициенты подобия по различным осям, так как размерности осей абсцисс и ординат имеют разный физический смысл [395]. Поэтому для сравнительного фрактального анализа молекул «клеточным» методом необходимо зафиксировать стандартную модель с начальной минимальной клеткой, размеры которой должны быть малы по сравнению с размахом гистограммы по обеим осям. Такой минимальной клетке соответствует значение $L=1$ в уравнении (155). Так как минимальная величина межатомного расстояния в молекуле составляет порядка $1\text{\AA}$, а точность расчета длин межатомных расстояний, полученная с помощью полуэмпирических квантово-химических методов, составляет $0.01\text{\AA}$ [405], то эта величина была выбрана в качестве минимально возможного размера клетки ΔR по оси абсцисс. В связи с тем, что минимальная, отличная от нуля, величина бара на оси ординат составляет 1, то в качестве разумного размера клетки Δf по оси ординат было установлено значение 0.01. Размер оцифрованной квадратной фрактальной матрицы определялся на основе максимума из двух величин: $R_{\max}/0.01$ и $f_{\max}/0.01$.

В качестве примера на рис. 26 и рис. 27 представлены результаты расчета фрактальной размерности гистограммы межатомных расстояний для молекулы метанола.

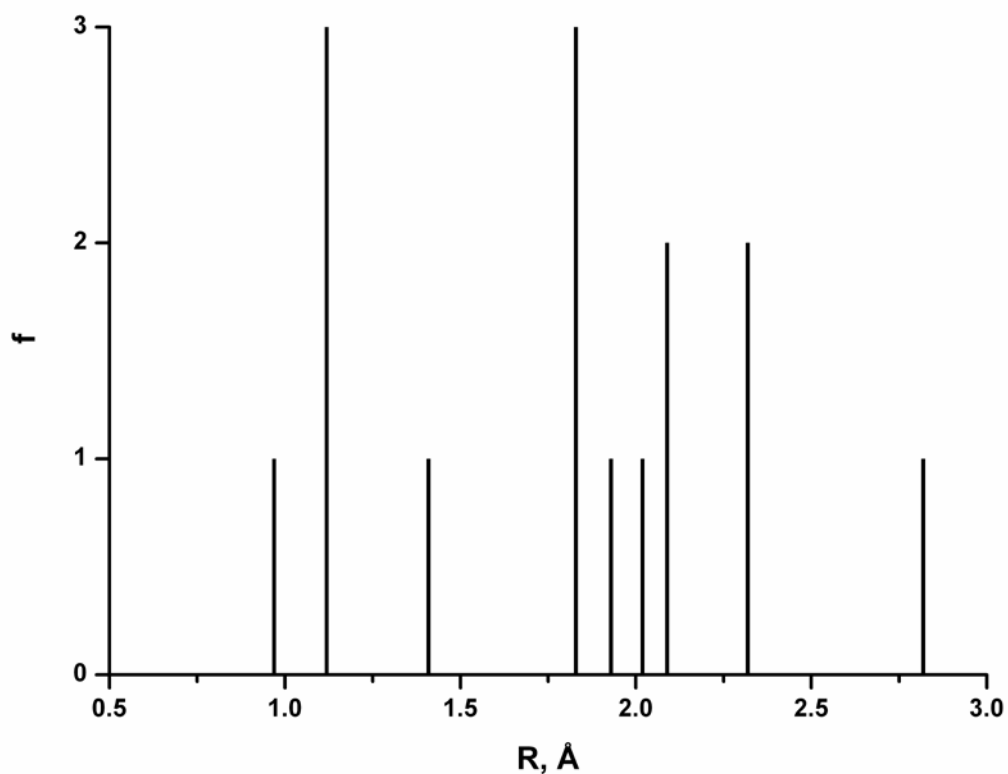


Рис. 26. Гистограмма межатомных расстояний в молекуле метанола, рассчитанная с разрешением 0.01 Å

Эта молекула содержит 6 атомов, поэтому общее число межатомных расстояний равно $6 \cdot (6-1)/2 = 15$, чему соответствует сумма всех баров на рис. 26. Для оцифровки гистограммы необходима фрактальная матрица размером 300 x 300 пикселей ($3/0.01=300$). В результате работы «клеточного» алгоритма были получены данные, представленные на рис. 27. Они описываются прямой линией с хорошими статистическими показателями. Полученное значение фрактальной размерности $D=1.06$.

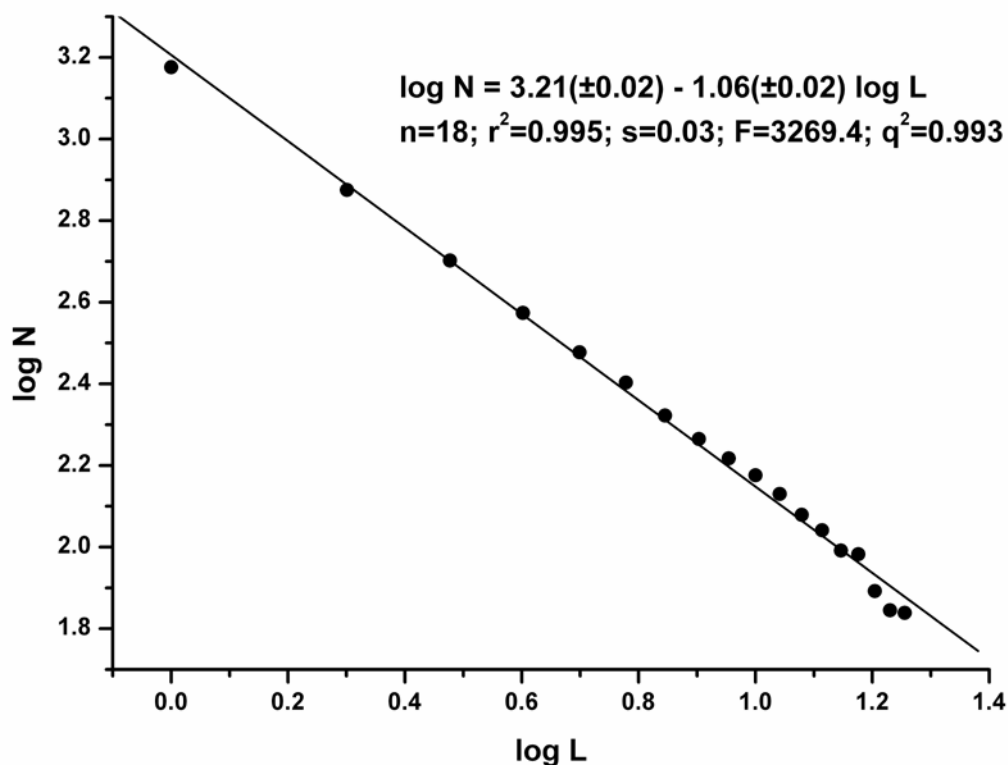


Рис. 27. Зависимость числа клеток (N) от их размера (L) (уравнение (155)) для молекулы метанола

В табл. 34 представлены результаты расчетов для 73 органических молекул, относящихся к 6 различным классам: алканы, спирты, эфиры, кетоны, амины, ароматические соединения. При этом число атомов N_a меняется от 5 до 50, а максимальная длина межатомных расстояний $R_{\max}$ варьирует в пределах от 1.82 Å до 20.57 Å. Минимальная величина фрактальной размерности D составляет 1.00, а максимальная – 1.58. Для сравнения: измерения фрактальной размерности, проведенные по описанной выше методике, дают для отрезка прямой значение $D=1$, а для квадрата $D=2$, что соответствует их топологическим размерностям.

Таблица 34

Точечные группы симметрии (S), число атомов (N_a), число баров в гистограмме межатомных расстояний (N_b), фрактальная размерность (D) молекул и статистические характеристики регрессионных уравнений

№	Молекула	S	N_a	N_b	D	n	r^2	s	F	q^2
1	Метан	T_d	5	2	1.00(± 0.01)	7	0.999	0.01	358483	0.999
2	Этан	D_{3d}	8	6	1.00(± 0.01)	19	0.999	0.01	307988	0.999
3	Пропан	C_{2v}	11	16	1.08(± 0.01)	32	0.998	0.02	17564	0.998
4	Бутан	C_{2h}	14	23	1.13(± 0.01)	44	0.997	0.02	16044	0.997
5	Пентан	C_{2v}	17	32	1.20(± 0.01)	57	0.992	0.04	6951	0.991
6	Гексан	C_{2h}	20	43	1.23(± 0.01)	69	0.992	0.04	7938	0.991
7	Гептан	C_{2v}	23	54	1.30(± 0.01)	82	0.991	0.05	9326	0.990
8	Октан	C_{2h}	26	64	1.31(± 0.01)	94	0.993	0.05	12260	0.992
9	Нонан	C_{2v}	29	76	1.33(± 0.01)	107	0.991	0.05	11748	0.990
10	Декан	C_{2h}	32	93	1.36(± 0.01)	119	0.992	0.05	15083	0.992
11	Ундекан	C_{2v}	35	96	1.36(± 0.01)	132	0.992	0.05	16576	0.991
12	Додекан	C_{2h}	38	106	1.37(± 0.01)	144	0.993	0.05	19326	0.992
13	Тридекан	C_{2v}	41	116	1.37(± 0.01)	157	0.993	0.05	22858	0.993
14	Тетрадекан	C_{2h}	44	132	1.37(± 0.01)	169	0.993	0.05	23095	0.992
15	Пентадекан	C_{2v}	47	139	1.38(± 0.01)	182	0.992	0.05	22773	0.992
16	Гексадекан	C_{2h}	50	153	1.39(± 0.01)	194	0.992	0.05	23793	0.991
17	Метанол	C_s	6	10	1.06(± 0.02)	18	0.995	0.03	3269	0.993
18	Этанол	C_s	9	23	1.22(± 0.02)	30	0.993	0.04	3777	0.991
19	1-Пропанол	C_s	12	33	1.23(± 0.01)	43	0.996	0.03	10460	0.995
20	2-Пропанол	C_1	12	39	1.21(± 0.01)	33	0.997	0.02	12298	0.994
21	1-Бутанол	C_s	15	49	1.25(± 0.01)	55	0.995	0.03	10609	0.994
22	2-Метил-1-пропанол	C_1	15	68	1.33(± 0.01)	43	0.998	0.02	20631	0.998
23	2-Бутанол	C_1	15	71	1.38(± 0.01)	46	0.996	0.03	10823	0.995
24	2-Метил-2-пропанол	C_s	15	36	1.20(± 0.01)	33	0.999	0.01	32528	0.999
25	1-Пентанол	C_s	18	60	1.27(± 0.01)	68	0.995	0.04	12865	0.994
26	1-Гексанол	C_s	21	80	1.32(± 0.01)	80	0.993	0.05	10518	0.992
27	1-Гептанол	C_s	24	98	1.36(± 0.01)	93	0.992	0.05	11146	0.991
28	1-Октанол	C_s	27	112	1.38(± 0.01)	105	0.993	0.05	13949	0.992
29	Диметиловый эфир	C_{2v}	9	13	1.03(± 0.01)	29	0.998	0.01	17938	0.998
30	Диэтиловый эфир	C_{2v}	15	34	1.26(± 0.02)	55	0.993	0.04	7545	0.992

Продолжение табл. 34

№	Молекула	S	N _a	N _b	D	n	r ²	s	F	q ²
31	Бутилметилловый эфир	C _s	18	70	1.38(±0.01)	67	0.994	0.04	10353	0.993
32	втор-Бутилметилловый эфир	C ₁	18	105	1.42(±0.01)	48	0.998	0.03	21181	0.997
33	трет-Бутилметилловый эфир	C _s	18	65	1.36(±0.01)	42	0.998	0.03	14145	0.997
34	Дипропиловый эфир	C _{2v}	21	59	1.36(±0.01)	79	0.992	0.05	9060	0.990
35	Диизопропиловый эфир	C ₂	21	83	1.41(±0.01)	55	0.995	0.04	11518	0.994
36	трет-Амилметилловый эфир	C ₁	21	138	1.44(±0.01)	47	0.998	0.02	25553	0.998
37	Бутилэтиловый эфир	C _s	21	86	1.40(±0.01)	80	0.995	0.04	16483	0.995
38	Дибутиловый эфир	C _{2v}	27	86	1.41(±0.01)	104	0.994	0.05	16003	0.993
39	Дипентилловый эфир	C _{2v}	33	125	1.45(±0.01)	129	0.992	0.05	15857	0.991
40	Диизопентилловый эфир	C ₂	33	202	1.52(±0.01)	104	0.997	0.03	33317	0.996
41	Ацетон	C _s	10	20	1.12(±0.01)	30	0.997	0.02	10729	0.997
42	2-Бутанон	C _s	13	40	1.24(±0.01)	42	0.998	0.02	18576	0.998
43	3-Метил-2-бутанон	C ₁	16	78	1.38(±0.01)	43	0.997	0.03	14187	0.996
44	2-Пентанон	C _s	16	58	1.28(±0.01)	57	0.995	0.04	11109	0.995
45	3-Пентанон	C _s	16	35	1.25(±0.01)	57	0.993	0.04	7646	0.992
46	2-Гексанон	C _s	19	67	1.31(±0.01)	69	0.995	0.04	14486	0.995
47	3-Гексанон	C _s	19	67	1.31(±0.01)	69	0.994	0.04	10935	0.993
48	2-Метил-3-пентанон	C ₁	19	110	1.38(±0.01)	57	0.996	0.03	15628	0.996
49	4-Метил-2-пентанон	C ₁	19	116	1.41(±0.01)	57	0.998	0.03	23218	0.997
50	Метиламин	C _s	7	12	1.14(±0.03)	19	0.990	0.04	1648	0.986
51	Диметиламин	C _s	10	22	1.15(±0.01)	31	0.996	0.03	6742	0.994
52	Этиламин	C ₁	10	35	1.30(±0.01)	31	0.997	0.03	9338	0.995
53	N-Этилметиламин	C ₁	13	56	1.37(±0.01)	44	0.996	0.03	10269	0.995
54	Изопропиламин	C ₁	13	53	1.32(±0.01)	33	0.997	0.03	10903	0.996
55	Пропиламин	C ₁	13	55	1.35(±0.01)	40	0.996	0.03	9001	0.995
56	Триметиламин	C _{3v}	13	15	1.08(±0.01)	30	0.996	0.03	6699	0.995
57	Бутиламин	C ₁	16	76	1.39(±0.01)	56	0.994	0.04	9355	0.993
58	втор-Бутиламин	C ₁	16	80	1.42(±0.01)	46	0.998	0.03	20548	0.997
59	трет-Бутиламин	C _s	16	43	1.27±0.01)	33	0.998	0.02	14740	0.998
60	Диэтиламин	C _s	16	54	1.39(±0.02)	57	0.993	0.05	7333	0.991
61	N,N-Диметилэтиламин	C ₁	16	81	1.36(±0.01)	43	0.997	0.03	15907	0.997
62	Изобутиламин	C ₁	16	80	1.37(±0.01)	44	0.998	0.02	21732	0.998
63	N-Метилизопропиламин	C ₁	16	87	1.38(±0.01)	44	0.998	0.02	26579	0.998
64	N-Метилпропиламин	C ₁	16	78	1.36(±0.01)	57	0.996	0.03	14017	0.996
65	Бензол	D _{6h}	12	10	1.07(±0.01)	38	0.995	0.03	6572	0.994

Продолжение табл. 34

№	Молекула	S	N _a	N _b	D	n	r ²	s	F	q ²
66	Нафталин	D _{2h}	18	39	1.31(±0.01)	61	0.993	0.04	8584	0.992
67	Антрацен	D _{2h}	24	66	1.41(±0.01)	84	0.992	0.05	10382	0.991
68	Фенантрен	C _{2v}	24	113	1.48(±0.01)	81	0.996	0.04	18046	0.995
69	2,3-Бензантрацен	D _{2h}	30	99	1.48(±0.01)	108	0.993	0.05	14368	0.992
70	1,2-Бензантрацен	C _s	30	252	1.52(±0.01)	104	0.997	0.03	39943	0.997
71	Пентацен	D _{2h}	36	133	1.50(±0.01)	132	0.994	0.05	22053	0.993
72	1,2:3,4-Дибензантрацен	C _{2v}	36	199	1.51(±0.01)	103	0.997	0.04	29468	0.996
73	1,2:5,6-Дибензантрацен	C _{2h}	36	235	1.58(±0.01)	124	0.996	0.04	30875	0.996

Данные, представленные на рис. 28, свидетельствуют о существовании тенденции между ростом числа атомов в молекуле и фрактальной размерностью. Однако, наблюдающиеся значительные отклонения от тренда, говорят о том, что число атомов не является единственным фактором. Например, четыре изомерных спирта с общей брутто-формулой C₄H₁₀O: 1-Бутанол(№ 21 в табл. 34), 2-Метил-1-пропанол(№ 22) , 2-Бутанол(№ 23) и 2-Метил-2-пропанол(№ 24) имеют по 15 атомов, но различаются по фрактальной размерности. Величина D при этом меняется от 1.20 до 1.38.

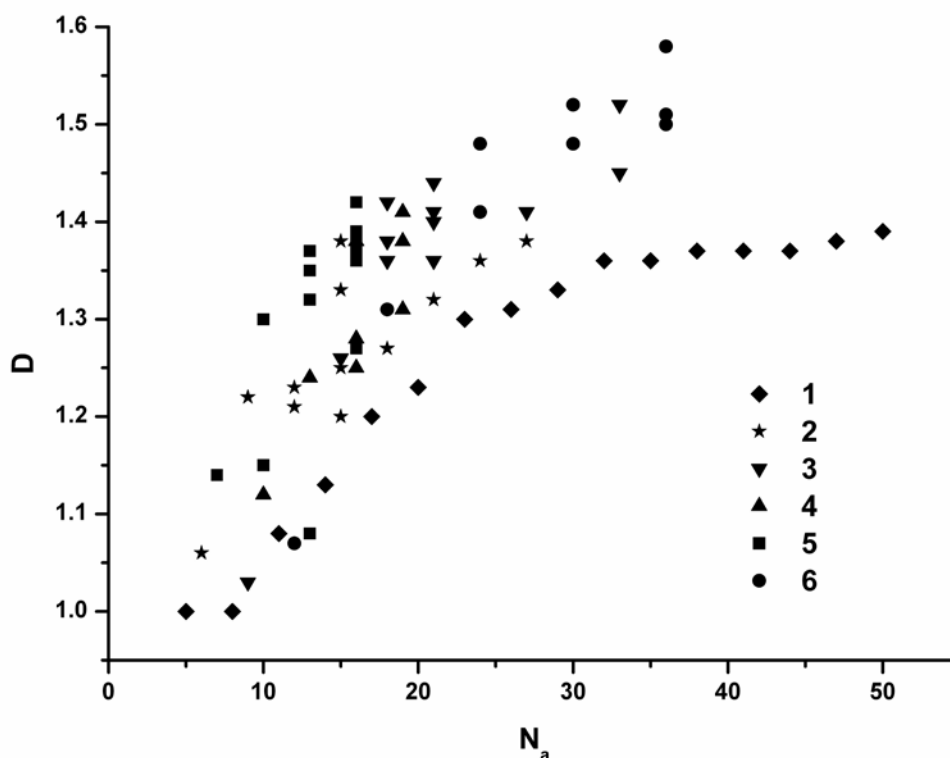


Рис. 28. Зависимость фрактальной размерности от числа атомов для различных молекул (1 – алканы, 2 – спирты, 3 – эфиры, 4 – кетоны, 5 – амины, 6 – ароматические соединения)

Наблюдающееся явление может быть объяснено, если принять во внимание точечные группы симметрии молекул. Действительно, соединения № 21 и № 24 относятся к группе C_s и имеют $D=1.25$ и 1.20 соответственно. Напротив, соединения № 22 и № 23 принадлежат к группе C_1 и их фрактальные размерности равны 1.33 и 1.38 соответственно. Количественной мерой, отражающей количество атомов в молекуле и принадлежность молекул к той или иной точечной группе симметрии, в первом приближении, может являться число баров в гистограммах межатомных расстояний N_b . Для 1-Бутанол $N_b=49$, для 2-Метил-2-пропанол – 36, для 2-Метил-1-пропанол – 68 и для 2-Бутанол – 71. Очевидно, что чем большим числом элементов симметрии обладает

молекула, тем больше будет число совпадающих межатомных расстояний и тем меньше будет представлено в гистограммах различных баров. Как показало, однако, дальнейшее исследование, более лучшей количественной мерой для оценки влияния числа атомов и точечной группы симметрии в молекуле является нормированная величина, получаемая путем деления числа баров на число атомов: N_b/N_a .

Регрессионный анализ зависимостей для отдельных точечных групп симметрии привел к следующим результатам:

$$C_{2v}: \quad D = 0.97(\pm 0.03) + 0.81(\pm 0.06) \log(N_b/N_a) \quad (156)$$

$n=14; r^2=0.935; s=0.04; F=172; q^2=0.889$

$$C_s: \quad D = 0.91(\pm 0.03) + 0.70(\pm 0.06) \log(N_b/N_a) \quad (157)$$

$n=23; r^2=0.851; s=0.04; F=120; q^2=0.810$

$$C_1: \quad D = 0.91(\pm 0.06) + 0.63(\pm 0.08) \log(N_b/N_a) \quad (158)$$

$n=18; r^2=0.785; s=0.03; F=58.5; q^2=0.682$

Уравнения (156)-(158) свидетельствуют о существовании корреляционной связи между фрактальной размерностью и нормированным числом баров. Эта связь становится понятной, если проанализировать физический смысл величины N_b/N_a . По сути, она представляет собой число баров, которое приходится на один атом, т.е. характеризует плотность заполнения пространства гистограммы. Поэтому фрактальная размерность, являясь количественной мерой степени изрезанности или сложности пространственной структуры молекул (в пределе, крайняя изрезанность – это полностью заполненное пространство), оказывается в статистической связи с величиной N_b/N_a .

Чтобы определить место фрактальной размерности гистограмм межатомных расстояний среди других дескрипторов, нами были проанализированы абсолютные значения парных коэффициентов линейной корреляции 3224 дескрипторов, рассчитанных с помощью программы DRAGON [315], с величиной D для 73 изученных соединений. Из указанного набора дескрипторов предварительно были удалены малоинформативные переменные, которые имели свыше 95% постоянных значений. В результате такого скрининга осталось 1363 дескриптора. На рис. 29 приведена гистограмма коэффициентов корреляции, рассчитанная с шагом 0.05. При этом минимальное значение $|r|$ равно 0.002, максимальное составляет 0.859, а среднее величина равна 0.332.

В работе [406] представлена классификация молекулярных дескрипторов в зависимости от их коэффициентов корреляции $|r|$ по отношению к газохроматографическому времени удерживания. В соответствии с этими данными, дескрипторы с $|r| \geq 0.99$ являются фундаментальными, $0.99 > |r| \geq 0.80$ – важными, $0.80 > |r| \geq 0.50$ – вероятными, $0.50 > |r|$ – специфичными. Предложенную классификацию можно использовать не только для характеристики дескрипторов по отношению к рассматриваемому свойству, но и рассматриваемого свойства, в частности, фрактальной размерности, по отношению к дескрипторам. Действительно, из данных, представленных на рис. 29, следует, что из 1363 дескрипторов 1010 имеют значение $|r| < 0.5$, 351 находятся в пределах $0.80 > |r| \geq 0.50$ и 2 дескриптора имеют $0.99 > |r| \geq 0.80$.

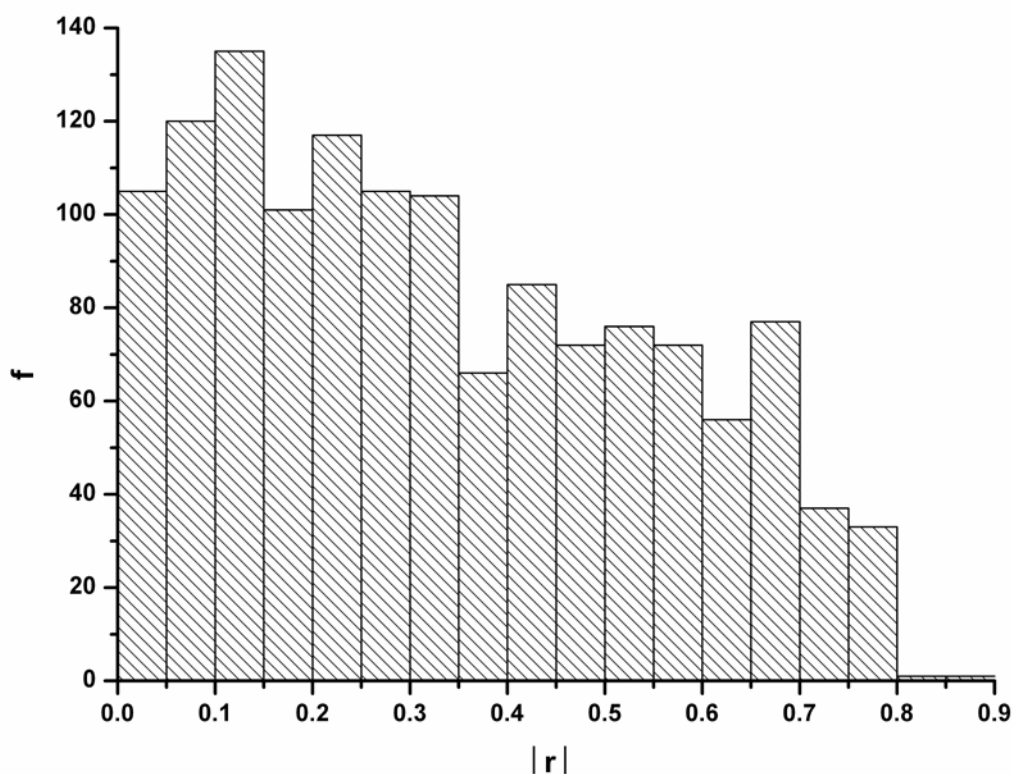


Рис. 29. Гистограмма распределения коэффициентов корреляции фрактальной размерности с 1363 дескрипторами

Интересно отметить, что два последних дескриптора представляют собой топологические индексы: структурное информационное содержание 1 и 2 порядка. Доля «специфичности» во фрактальной размерности по отношению к 1363 дескрипторам составляет $100 \cdot 1010 / 1363 = 74.1\%$, доля «вероятности» $100 \cdot 351 / 1363 = 25.8\%$, а оставшаяся доля «важности» 0.1% . Т.е. фрактальную размерность можно рассматривать в качестве весьма «специфического» дескриптора.

Полезность нового предлагаемого дескриптора может быть продемонстрирована на примере установления количественных соотношений (уравнение (159)) между фрактальной размерностью и плотностью для гомологического ряда алканов (пентан, гексан, ...эйкозан, пентакозан,

триаконтан), которые при нормальных условиях представляют собой жидкости или твердые тела (табл. 35).

$$d_4^{20} \text{ (г/см}^3\text{)} = -0.21(\pm 0.05) + 0.70(\pm 0.03) D \quad (159)$$

$$n=18; r^2=0.965; s=0.01; F=442; q^2=0.956$$

Полученное регрессионное уравнение свидетельствует о существовании тесной корреляционной связи между соответствующими микро- и макромолекулярными характеристиками. В изученном ряду соединений возрастание сложности молекулярной структуры сопровождается увеличением плотности вещества. Таким образом, фрактальная размерность может быть использована в качестве дескриптора при исследовании связи «структура – свойство» и «структура – активность».

В заключение следует отметить, что предлагаемый дескриптор вполне удовлетворяет основным требованиям, предъявляемым к новым дескрипторам [407]. Он является инвариантным по отношению к обозначению и нумерации молекулярных атомов. Величина D не зависит от молекулярного вращения и трансляции, так как в основе ее расчета лежит инвариант: межатомное расстояние. Приведенное описание расчета фрактальной размерности гистограммы межатомных расстояний не является двусмысленным. И, наконец, диапазон изменения D от 1 до 2 является удобным с практической точки зрения.

Таблица 35

Величины фрактальной размерности (D) и плотности (d_4^{20}) в гомологическом ряду алканов

№	Алкан	D	d_4^{20} , г/см ³
1	Пентан	1.20	0.6262
2	Гексан	1.23	0.6603
3	Гептан	1.30	0.6838
4	Октан	1.31	0.7025
5	Нонан	1.33	0.7176
6	Декан	1.36	0.7300
7	Ундекан	1.36	0.7402
8	Додекан	1.37	0.7487
9	Тридекан	1.37	0.7564
10	Тетрадекан	1.37	0.7628
11	Пентадекан	1.38	0.7685
12	Гексадекан	1.39	0.7730
13	Гептадекан	1.39	0.7780
14	Октадекан	1.41	0.7820
15	Нонадекан	1.42	0.7855
16	Эйкозан	1.42	0.7886
17	Пентакозан	1.46	0.8012
18	Триаконтан	1.47	0.8097

6.3. Линейная динамическая модель

При исследовании количественной связи «структура – активность» и «структура – свойство» одной из основных целей является конструирование предсказательных моделей биологической активности и физико-химических свойств как функции молекулярной и структурной информации в рядах химических соединений [287, 408]. В настоящее время широкое распространение для изучения таких связей получили такие методы, как множественный регрессионный анализ, проекции на латентные структуры, нейронные сети и метод опорных векторов [409]. Общей характерной чертой рассматриваемых подходов является то, что в качестве независимых переменных выступают различные физико-химические, электронные, топологические дескрипторы, представленные в виде матрицы, а в качестве зависимой переменной – вектор активностей (свойств). При геометрической интерпретации каждая молекула представляется точкой в многомерном пространстве дескрипторов. При этом точечные значения для некоторых экспериментальных дескрипторов извлекаются из соответствующих двумерных кривых. Например, нормальная точка кипения вещества присутствует в кривой зависимости давления насыщенного пара от температуры, а величина LD₅₀ берется из графика «доза – эффект».

Известно [1], что кривая «доза – эффект» отражает свойства как вещества, так и организма, на который оно действует, т.е. содержит информацию о механизме действия. Очевидно, что количество информации, содержащейся в двумерной кривой, больше, чем в отдельной точке. Поэтому представляет интерес исследование количественной связи между структурой и свойствами молекул, выраженных в виде двумерных кривых. Одной из форм представления структуры молекул в QSAR/QSPR исследованиях является функция радиального распределения, предложенная в работе [322] и нашедшая дальнейшее применение [410-412]. А в качестве одного из подходов для

установления количественных закономерностей может быть использована линейная динамическая модель [413].

В настоящей работы проведено исследование количественной связи между структурой молекул, выраженной в виде функции радиального распределения, и кривой зависимости давления насыщенного пара от температуры в ряду *n*-алканов на основе линейной динамической модели.

Функцию радиального распределения рассчитывали на основе математического аппарата газовой электронографии, модифицированного для использования в QSAR/QSPR исследованиях [322]:

$$f(R) = \frac{2}{\pi} \int_0^{s_{\max}} s M(s) \exp(-bs^2) \sin sR ds \quad (160)$$

здесь $f(R)$ - функция радиального распределения, R - межатомное расстояние, $M(s)$ - функция молекулярного рассеяния, s - угловой параметр.

$$s = \frac{4\pi}{\lambda} \sin \frac{\theta}{2} \quad (161)$$

λ – длина волны, θ – угол рассеяния, b – параметр, оцениваемый из уравнения:

$$\exp(-bs_{\max}^2) = 0.1 \quad (162)$$

В случае жесткой модели молекулы (отсутствие внутримолекулярных колебаний):

$$M(s) = \sum_{i=1}^N \sum_{j>i}^N C_{ij} \frac{\sin sR_{ij}}{sR_{ij}} \quad (163)$$

здесь N – число атомов в молекуле, C_{ij} – коэффициент, характеризующий рассеивающую способность данной пары атомов i и j :

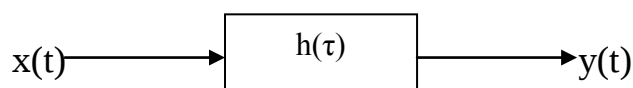
$$C_{ij} = \frac{Z_i Z_j}{\sum_{k=1}^N (Z_k^2 + Z_k)} \quad (164)$$

здесь Z – порядковый номер атома.

При моделировании в качестве независимой переменной выступало межатомное расстояние R , а в качестве зависимой – функция радиального распределения $f(R)$. При этом минимальное значение R было равно 0Å , а максимальное значение 16Å соответствовало самому большому межатомному расстоянию в молекуле додекана (15.57Å).

Данные по зависимости давления насыщенного пара (P) 10 н-алканов (пропан, бутан,...додекан) от температуры (T) (в форме уравнений Антуана) были взяты из работы [309]. P менялось в пределах от 10 мм Hg до 1500 мм Hg, а T варьировало в пределах от 164 К до 520 К. Учитывая, что давление насыщенного пара имело одинаковую величину интервала для всех рассматриваемых углеводородов, оно было выбрано в качестве независимой переменной, а температуры кипения – в качестве моделируемой функции.

Количественную связь между исследуемыми кривыми устанавливали на основе линейной динамической модели с одним входом $x(t)$, одним выходом $y(t)$ и импульсной переходной функцией $h(\tau)$ [413]:



Связь между процессами $x(t)$ и $y(t)$ может быть выражена следующим образом:

$$y(t) = \int_0^{\infty} h(\tau)x(t-\tau)d\tau \quad (165)$$

При переходе из временной (физической) области в частотную может быть получено следующее уравнение:

$$Y(f) = H(f) X(f) \quad (166)$$

здесь $X(f)$ и $Y(f)$ – преобразования Фурье для $x(t)$ и $y(t)$ соответственно, а $H(f)$ – частотная характеристика. Частотную характеристику вычисляли путем усреднения по всем кривым на основе соотношения для взаимного спектра входного и выходного процессов:

$$S_{xy}(f) = H(f) S_{xx}(f) \quad (167)$$

здесь $S_{xy}(f)$ – взаимная двусторонняя спектральная плотность, $S_{xx}(f)$ – двусторонняя спектральная плотность $x(t)$.

Качество исследуемой модели оценивали с помощью функции когерентности, изменяющейся в пределах от 0 до 1 и вычисляемой по формуле:

$$\gamma_{xy}^2(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \quad (168)$$

где $S_{yy}(f)$ – двусторонняя спектральная плотность $y(t)$.

В дальнейшем для сравнения различных моделей использовали среднее значение γ^2 , полученное путем усреднения функции когерентности по всем частотам.

Расчеты выполняли с использованием модифицированной компьютерной программы на основе быстрого преобразования Фурье [414]. В качестве входных и/или выходных процессов выступали $f(R)$ и обратная функция давления насыщенного пара от температуры $T=f(P)$. Оцифровку зависимостей

проводили путем дискретизации независимых переменных с использованием 16, 32, 64, 128, 256 точек (уровни дискретизации m) и вычислением соответствующих значений функций.

Учитывая разный физический смысл входных и выходных кривых, для получения статистически значимых корреляционных зависимостей использовали модификацию линейной динамической модели. Для этого на вход системы подавали участки кривых различной длины. При использовании $f(R)$ в качестве входной кривой (модель «структура – активность») проводили варьирование ее длины путем деления интервала R ($0 - 16 \text{ \AA}$) на 160 частей и парной комбинации полученных значений $R_{\text{мин}}$ и $R_{\text{макс}}$. При этом длина выходной кривой $T=f(P)$ оставалась неизменной и была максимальной ($P_{\text{мин}}=10 \text{ мм Hg}$, $P_{\text{макс}}=1500 \text{ мм Hg}$). Аналогичный подход использовали для входной кривой $T=f(P)$ (модель «активность – структура»). В этом случае интервал P ($10 - 1500 \text{ мм Hg}$) делили на 50 частей и комбинировали величины $P_{\text{мин}}$ и $P_{\text{макс}}$. Длина кривой $f(R)$ была постоянной ($R_{\text{мин}}=0 \text{ \AA}$, $R_{\text{макс}}=16 \text{ \AA}$). Полученные модели сортировали по величине остаточного стандартного отклонения и отбирали те из них, которые имели минимальное значение.

На рис. 30 представлен пример функции $f(R)$ для молекул пропана и додекана.

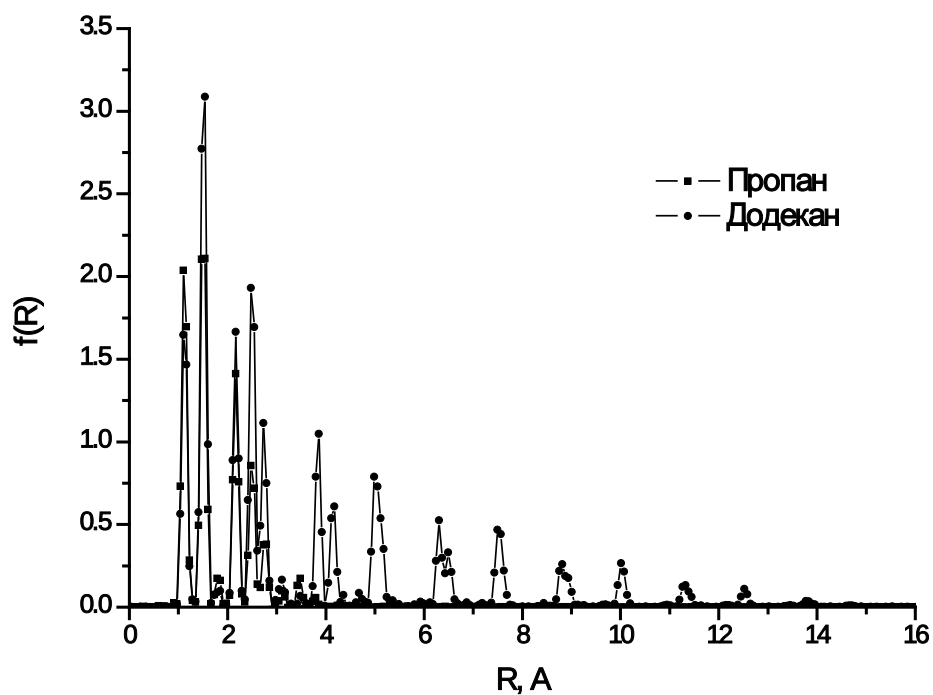


Рис. 30. Функция радиального распределения для молекул пропана и додекана при уровне дискретизации $m=256$

Максимумы на кривых $f(R)$ соответствуют парным межатомным расстояниям. Так, первый пик, наблюдаемый в районе $R=1.10 \text{ \AA}$, обусловлен наличием в молекулах валентно связанных атомов C-H, а второй пик при $R=1.54 \text{ \AA}$ соответствует валентно связанным атомам C-C. Последующие максимумы отражают расстояния между несвязанными атомами.

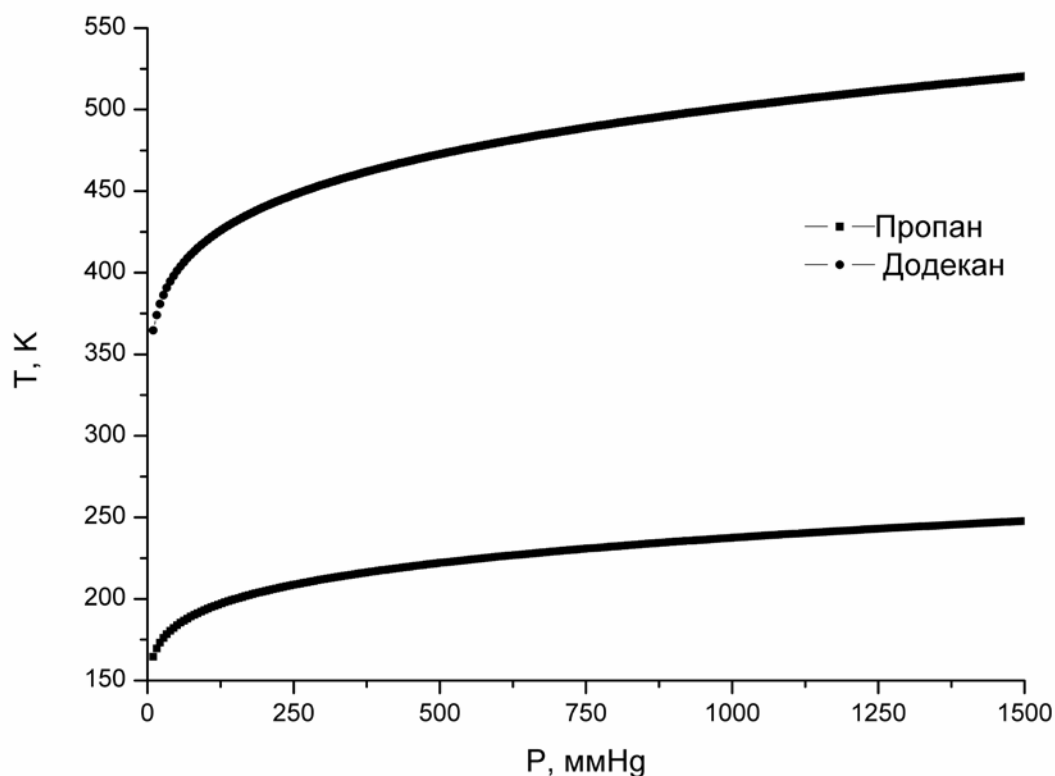


Рис. 31. Обратная функция зависимости давления насыщенного пара от температуры для молекул пропана и додекана при уровне дискретизации $m=256$

Представленные на рис. 31 кривые $T=f(P)$ являются типичными для рассматриваемого ряда n -алканов. Они имеют одинаковый интервал изменения по оси абсцисс (10...1500 мм Hg), но отличаются значениями по оси ординат. При этом температура меняется от 164 К до 248 К (пропан) и от 364 К до 520 К (додекан).

В табл. 36 приведены характеристики лучших моделей типа «структура – активность», полученных путем перебора 12880 ($161 \cdot 160/2$) комбинаций различной длины на каждом уровне дискретизации.

Таблица 36

Порядковый номер модели (№), уровни дискретизации (m), минимальная ($R_{\text{мин}}$) и максимальная ($R_{\text{макс}}$) величина межатомного расстояния, остаточное стандартное отклонение (s) и среднее значение функции когерентности (γ^2) для моделей «структура – активность»

№	m	$R_{\text{мин}}, \text{\AA}$	$R_{\text{макс}}, \text{\AA}$	s	γ^2
1	16	1.1	13.9	5.0	0.988
2	32	0.8	12.7	5.4	0.944
3	64	1.4	15.3	5.2	0.976
4	128	0.0	15.5	6.6	0.956
5	256	0.0	15.6	6.6	0.933

Из данных таблицы следует, что при увеличении степени дискретизации величина s имеет тенденцию к увеличению. Среднее значение функции когерентности γ^2 ведет себя противоположным образом. Длина кривой $f(R)$ принимает разные значения. При этом $R_{\text{мин}}$ меняется от 0 $\text{\AA}$ до 1.4 $\text{\AA}$, а $R_{\text{макс}}$ меняется от 12.7 $\text{\AA}$ до 15.6 $\text{\AA}$. Учитывая, что модель 1 имеет минимальную величину $s=5.0$ и максимальное значение $\gamma^2=0.988$, она была выбрана в качестве финальной модели типа «структура-активность».

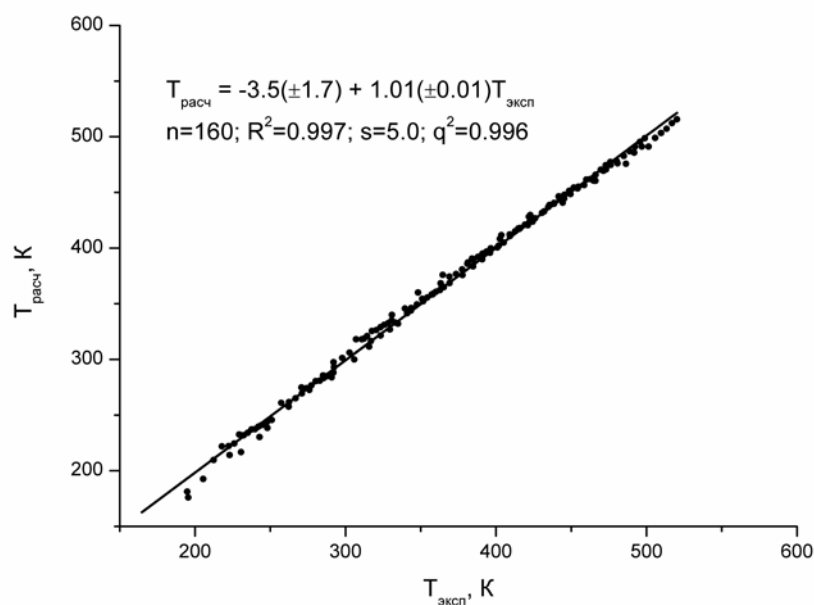


Рис. 32. Зависимость рассчитанных по модели 1 значений температур кипения от экспериментальных величин

Данные, представленные на рис. 32, свидетельствуют о том, что модель 1 адекватно описывает кривую $T=f(P)$. Об этом говорят как высокие значения характеристик описательной ($r^2=0.997$, $s=5.0$) и предсказательной ($q^2=0.996$) способности модели, так и величины коэффициентов.

Следует отметить, что при проведении QSAR/QSPR исследований кроме задачи получения адекватных моделей, связывающих структуру молекул с их активностью и свойствами («прямая» задача), стоят и другие задачи. В качестве своеобразной «сверхзадачи» можно рассматривать предсказание структур с заданными свойствами («обратная» задача). Этой теме уделяется большое внимание в литературе [415-421]. Одним из возможных путей в решении этой проблемы является получение качественных QSAR/QSPR моделей и компьютерная генерация на их основе молекулярных структур с учетом различных ограничений: физических, химических и др.

В рамках рассматриваемой модифицированной линейной динамической системы эта «сверхзадача» решается естественным образом. Используя уравнение (166) можно при знании $H(f)$ и $Y(f)$ оценивать величину $X(f)$ и наоборот: комбинация известных величин $H(f)$ и $X(f)$ позволяет вычислить $Y(f)$. Путем обратного преобразования Фурье можно вернуться к исходной физической системе. Например, используя модель 1 (табл. 36) и задавая кривую $T=f(P)$, можно получить $f(R)$. Однако, качество получаемой функции радиального распределения оказывается слишком плохим, чтобы на ее основе делать надежные умозаключения о трехмерной структуре молекул. В самом деле, разрешение по оси ординат (в соответствии с характеристиками модели 1) составляет всего $(13.9-1.1)/(16-1)=0.85 \text{ \AA}$, а $R_{\text{макс}}=13.9 \text{ \AA}$, что явно недостаточно для получения надежной структурной информации. Модели 2-5 имеют более высокую разрешающую способность, однако они не охватывают весь диапазон изменения межатомных расстояний. Поэтому мы сконструировали специальную модель «активность - структура», где в качестве входа используется зависимость $T=f(P)$ переменной длины, а в качестве выхода – полноразмерная $f(R)$.

Таблица 37

Порядковый номер модели (№), уровни дискретизации (m), минимальная ($P_{\text{мин}}$) и максимальная ($P_{\text{макс}}$) величина давления насыщенного пара, остаточное стандартное отклонение (s) и среднее значение функции когерентности (γ^2) для моделей «активность - структура»

№	m	$P_{\text{мин}}$, мм Hg	$P_{\text{макс}}$, мм Hg	s	γ^2
6	16	1470.2	1500.0	0.018	0.992
7	32	10.0	39.8	0.047	0.893
8	64	1470.2	1500.0	0.072	0.919
9	128	1470.2	1500.0	0.059	0.943
10	256	1470.2	1500.0	0.061	0.940

Данные, представленные в табл. 37, свидетельствуют о том, что наиболее «информативным» участком кривой $T=f(P)$ является отрезок 1470.2 – 1500 мм Hg. При увеличении степени дискретизации s увеличивается с 0.018 до 0.061, а γ^2 уменьшается с 0.992 до 0.940. Учитывая, что точность описания функции радиального распределения по оси ординат (интенсивность) на уровне величины 0.06 представляется приемлемой, а максимальное разрешение по оси абсцисс (межатомное расстояние) достигается при $m=256$ и составляет $(16-0)/(256-1)=0.06 \text{ \AA}$, в качестве рабочей модели «активность – структура» была использована зависимость 10. Используя эту модель и схему кросс-валидации с выбором по одному, были рассчитаны величины $f(R)$ для всех исследованных n-алканов. Ниже представлено соответствующее регрессионное уравнение:

$$f(R)_{\text{расч}} = -0.001(\pm 0.001) + 1.00(\pm 0.01) f(R)_{\text{эксп}} \quad (169)$$

$n=2560; r^2=0.971; s=0.061; q^2=0.966$

Параметры и коэффициенты регрессионного уравнения (169) свидетельствуют о высоком качестве модели 10, которая может быть применена для вычисления $f(R)$ в исследованном ряду n-алканов. Это

демонстрирует и представленный на рис. 33 пример сопоставления рассчитанных и экспериментальных значений функции радиального распределения для молекулы гептана с хорошим совпадением величин как по абсциссе, так и по ординате.

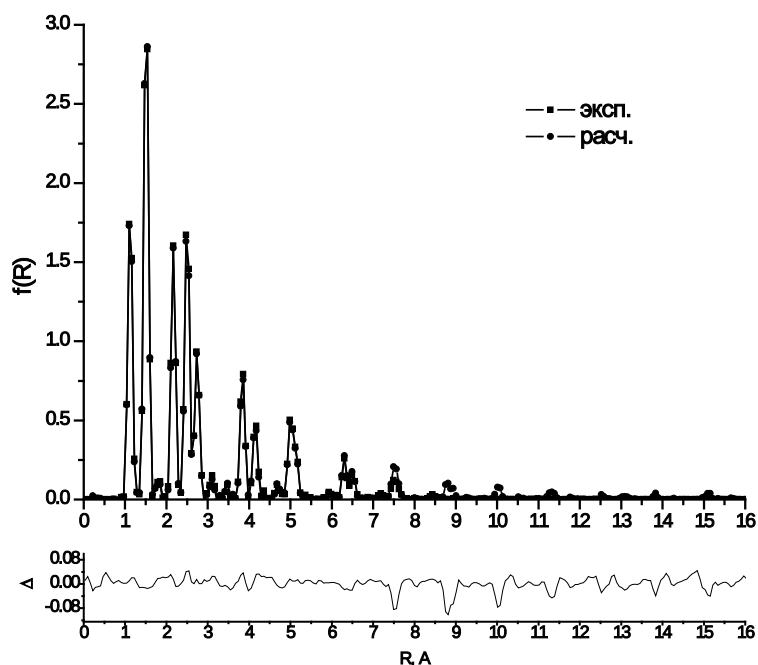


Рис. 33. Экспериментальные и рассчитанные по модели 10 значения функции радиального распределения и их разность (Δ) для молекулы гептана при уровне дискретизации $m=256$

Таким образом, «обратная» задача (предсказание структур с заданными свойствами) в рамках модифицированной линейной динамической модели может быть решена путем конструирования модели типа «активность – структура», расчета функции радиального распределения для молекулы с заданными свойствами и подбора и сравнения с $f(R)$ модельного соединения с известной трехмерной структурой.

Выводы по главе

Показано, что в качестве модели для описания острой токсичности в гомологических рядах органических соединений может быть использовано простое рекуррентное уравнение. При этом было выявлено, что рекуррентная модель токсичности обладает лучшей способностью к экстраполяции данных по сравнению с линейной и параболической моделями. При анализе коэффициентов рекуррентных уравнений с учетом типов химических соединений, удастся выявить неспецифический или специфический характер токсического действия гомологического ряда. Аналогичный анализ с использованием биологических объектов позволяет оценить их токсикологическую близость между собой с помощью расстояния между ними.

Предложенный 3D дескриптор молекулярной структуры: фрактальная размерность гистограмм межатомных расстояний, может служить мерой сложности структуры в ее геометрическом аспекте. Выявлено, что новый дескриптор зависит, главным образом, от трех факторов: числа атомов в молекуле, числа баров в гистограмме и точечной группы симметрии. Установлено, что фрактальная размерность гистограмм межатомных расстояний обладает весьма специфическим характером.

Разработанная модифицированная линейная динамическая модель, устанавливающая связь между двумерными кривыми, описывающими зависимость давления насыщенного пара от температуры и функциями радиального распределения ряда *n*-алканов, имеет хорошие статистические характеристики. Установлено, что эта модель может быть использована в качестве рабочего инструмента при проведении QSAR/QSPR исследований не только для получения корреляционных зависимостей между структурой и активностью, но также для извлечения информации о трехмерной структуре молекул с заданными свойствами.

Заключение

1. Разработана аддитивно-мультипликативная модель для описания энтальпии и свободной энергии Н-комплексов. На основе этой модели создана единая энергетическая шкала водородной связи и рассчитаны дескрипторы, количественно характеризующие способность соединений к образованию Н-связи.
2. Разработана количественная модель для описания липофильности органических соединений на основе поляризуемости и свободноэнергетических протоноакцепторных дескрипторов.
3. На базе стандартных и оригинальных подходов сконструированы и протестированы модели острой токсичности органических соединений по отношению к водным организмам и грызунам с применением дескрипторов водородной связи и других дескрипторов. Продемонстрирована их применимость для предсказания токсичности больших массивов органических соединений разнообразных химических классов.
4. С использованием физико-химических дескрипторов: молекулярной поляризуемости, свободноэнергетического протоноакцепторного дескриптора и стерического параметра разработаны модели антигиббереллиновой (ретардантной) активности солей четвертичного аммония.
5. Создана модель для предсказания проницаемости физиологически активных веществ через гематоэнцефалический барьер на основе величин структурного сходства и дескрипторов водородной связи.
6. Разработана рекуррентная модель токсичности в гомологических рядах соединений. Выявлено, что с помощью этой модели может быть установлен тип токсического действия ряда и токсикологическая близость биообъектов между собой.

7. Для описания структуры химических соединений создан и применен новый дескриптор: фрактальная размерность. Установлено, что предлагаемый дескриптор описывает сложность геометрической структуры молекул и может быть использован при исследовании связи «структура – свойство» и «структура – активность».
8. Разработан новый подход для установления количественной связи между структурой и свойствами (активностью) химических соединений на основе модифицированной линейной динамической модели, который может быть использован в качестве рабочего инструмента при проведении QSAR/QSPR исследований.
9. На основе теоретических разработок получен ряд практических результатов: создана база данных по термодинамическим характеристикам комплексов с водородной связью; сформирована база данных, содержащая Н-донорные и Н-акцепторные дескрипторы органических соединений; разработаны компьютерные программы, ориентированные на использование в QSAR/QSPR моделировании.

Список литературы

1. Куценко С.А. Основы токсикологии. СПб.: Фолиант, 2004. 720 с.
2. Schultz T.W., Cronin M.T.D., Walker J.D., Aptula A.O. Quantitative Structure–Activity Relationships (QSARs) in Toxicology: a Historical Perspective // J. Mol. Struct. (Theochem). 2003. Vol. 622. № 1-2. P. 1-22.
3. Schultz T.W., Cronin M.T.D., Netzeva T.I. The Present Status of QSAR in Toxicology // J. Mol. Struct. (Theochem). 2003. Vol. 622. № 1-2. P. 23-38.
4. Bradbury S.P., Russom C.L., Ankley G.T., Schultz T.W., Walker J.D. Overview of Data and Conceptual Approaches for Derivation of Quantitative Structure–Activity Relationships for Ecotoxicological Effects of Organic Chemicals // Environ. Toxicol. Chem. 2003. Vol. 22. № 8. P. 1789-1798.
5. Lessigiarska I., Worth A.P., Netzeva T.I. Comparative Review of (Q)SARs for Acute Toxicity // EUR 21559 EN. European Commission. Joint Research Centre. Ispra. Italy. 2005. 144 p.
6. Netzeva T.I., Pavan M., Worth A.P. Review of (Quantitative) Structure – Activity Relationships for Acute Aquatic Toxicity // QSAR Comb. Sci. 2008. Vol. 27. № 1. P. 77-90.
7. Devillers J., Devillers H. Prediction of Acute Mammalian Toxicity from QSARs and Interspecies Correlations // SAR QSAR Environ. Res. 2009. Vol. 20. № 5-6. P. 467-500.
8. Tsakovska I., Lessigiarska I., Netzeva T., Worth A. Review of (Q)SARs for Mammalian Toxicity // EUR 22486 EN. European Commission. Joint Research Centre. Ispra. Italy. 2006. 60 p.
9. Tsakovska I., Lessigiarska I., Netzeva T., Worth A. A Mini Review of Mammalian Toxicity (Q)SAR Models // QSAR Comb. Sci. 2008. Vol. 27. № 1. P. 41-48.
10. Dearden J.C. In Silico Prediction of Drug Toxicity // J. Comput. Aided Mol. Des. 2003. Vol. 17. № 2-4. P. 119-127.

11. Perkins R., Fang H., Tong W., Welsh W.J. Quantitative Structure-Activity Relationship Methods: Perspectives on Drug Discovery and Toxicology // *Environ. Toxicol. Chem.* 2003. Vol. 22. № 8. P. 1666-1679.
12. Kaiser K.L.E. The Use of Neural Networks in QSARs for Acute Aquatic Toxicological Endpoints // *J. Mol. Struct. (Theochem)*. 2003. Vol. 622. № 1-2. P. 85-95.
13. Kaiser K.L.E. Neural Networks for Effect Prediction in Environmental and Health Issues Using Large Datasets // *QSAR Comb. Sci.* 2003. Vol. 22. № 2. P. 185-190.
14. Escher B.I., Hermens J.L.M. Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects // *Environ. Sci. Technol.* 2002. Vol. 36. № 20. P. 4201-4217.
15. Selassie C.D., Garg R., Kapur S., Kurup A., Verma R.P., Mekapati S.B., Hansch C. Comparative QSAR and the Radical Toxicity of Various Functional Groups // *Chem. Rev.* 2002. Vol. 102. № 7. P. 2585-2605.
16. Katritzky A.R., Tatham D.B., Maran U. Theoretical Descriptors for the Correlation of Aquatic Toxicity of Environmental Pollutants by Quantitative Structure-Toxicity Relationships // *J. Chem. Inf. Comput. Sci.* 2001. Vol. 41. № 5. P. 1162-1176.
17. Cronin M.T.D., Netzeva T.I., Dearden J.C., Edwards R., Worgan A.D.P. Assessment and Modeling of the Toxicity of Organic Chemicals to *Chlorella vulgaris*: Development of a Novel Database // *Chem. Res. Toxicol.* 2004. Vol. 17. № 4. P. 545-554.
18. Worgan A.D.P., Dearden J.C., Edwards R., Netzeva T.I., Cronin M.T.D. Evaluation of a Novel Short-Term Algal Toxicity Assay by the Development of QSAR and Inter-Species Relationships for Narcotic Chemicals // *QSAR Comb. Sci.* 2003. Vol. 22. № 2. P. 204-209.

19. Ren S., Schultz T.W. Identifying the Mechanism of Aquatic Toxicity of Selected Compounds by Hydrophobicity and Electrophilicity Descriptors // *Toxicol. Lett.* 2002. Vol. 129. № 1-2. P. 151-160.
20. Ivanciuc O. Support Vector Machine Identification of the Aquatic Toxicity Mechanism of Organic Compounds // *Internet Electron. J. Mol. Des.* 2002. Vol. 1. № 3. P. 157–172.
21. Enoch S.J., Hewitt M., Cronin M.T.D., Azam S., Madden J.C. Classification of Chemicals According to Mechanism of Aquatic Toxicity: An Evaluation of the Implementation of the Verhaar Scheme in Toxtree // *Chemosphere*. 2008. Vol. 73. № 3. P. 243-248.
22. Seward J.R., Hamblen E.L., Schultz T.W. Regression Comparisons of *Tetrahymena pyriformis* and *Poecilia Reticulata* Toxicity // *Chemosphere*. 2002. Vol. 47. № 1. P. 93-101.
23. Cronin M.T.D., Aptula A.O., Duffy J.C., Netzeva T.I., Rowe P.H., Valkova I.V., Schultz T.W. Comparative Assessment of Methods to Develop QSARs for the Prediction of the Toxicity of Phenols to *Tetrahymena pyriformis* // *Chemosphere*. 2002. Vol. 49. № 10. P. 1201-1221.
24. González M.P., Díaz H.G., Cabrera M.A., Ruiz R.M. A Novel Approach to Predict a Toxicological Property of Aromatic Compounds in the *Tetrahymena pyriformis* // *Bioorg. Med. Chem.* 2004. Vol. 12. № 4. P. 735-744.
25. Schultz T.W., Netzeva T.I., Cronin M.T.D. Selection of Data Sets for QSARs: Analyses of *Tetrahymena* Toxicity from Aromatic Compounds // *SAR QSAR Environ. Res.* 2003. Vol. 14. № 1. P. 59-81.
26. Schultz T.W., Netzeva T.I., Roberts D.W., Cronin M.T.D. Structure-Toxicity Relationships for the Effects to *Tetrahymena pyriformis* of Aliphatic, Carbonyl-Containing, α,β -Unsaturated Chemicals // *Chem. Res. Toxicol.* 2005. Vol. 18. № 2. P. 330-341.
27. Cronin M.T.D., Schultz T.W. Development of Quantitative Structure-Activity Relationships for the Toxicity of Aromatic Compounds to *Tetrahymena*

- pyriformis: Comparative Assessment of the Methodologies // Chem. Res. Toxicol. 2001. Vol. 14. № 9. P. 1284-1295.
28. Schultz T.W., Cronin M.T.D., Netzeva T.I., Aptula A.O. Structure-Toxicity Relationships for Aliphatic Chemicals Evaluated with *Tetrahymena pyriformis* // Chem. Res. Toxicol. 2002. Vol. 15. № 12. P. 1602-1609.
29. Koleva Y.K., Madden J.C., Cronin M.T.D. Formation of Categories from Structure-Activity Relationships To Allow Read-Across for Risk Assessment: Toxicity of α,β -Unsaturated Carbonyl Compounds // Chem. Res. Toxicol. 2008. Vol. 21. № 12. P. 2300-2312.
30. Schultz T.W., Seward J.R. Dimyristoyl Phosphatidylcholine/Water Partitioning-Dependent Modeling of Narcotic Toxicity to *Tetrahymena pyriformis* // Quant. Struct.-Act. Relat. 2000. Vol. 19. № 4. P. 339-344.
31. Schultz T.W., Hewitt M., Netzeva T.I., Cronin M.T.D. Assessing Applicability Domains of Toxicological QSARs: Definition, Confidence in Predicted Values, and the Role of Mechanisms of Action // QSAR Comb. Sci. 2007. Vol. 26. № 2. P. 238-254.
32. Cronin M.T.D., Manga N., Seward J.R., Sinks G.D., Schultz T.W. Parametrization of Electrophilicity for the Prediction of the Toxicity of Aromatic Compounds // Chem. Res. Toxicol. 2001. Vol. 14. № 11. P. 1498-1505.
33. Kiralj R., Ferreira M.M.C. Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application // J. Braz. Chem. Soc. 2009. Vol. 20. № 4. P. 770-787.
34. Yao X.J., Panaye A., Doucet J.P., Zhang R.S., Chen H.F., Liu M.C., Hu Z.D., Fan B.T. Comparative Study of QSAR/QSPR Correlations Using Support Vector Machines, Radial Basis Function Neural Networks, and Multiple Linear Regression // J. Chem. Inf. Comput. Sci. 2004. Vol. 44. № 4. P. 1257-1266.
35. Aptula A.O., Netzeva T.I., Valkova I.V., Cronin M.T.D., Schultz T.W., Kuhne R., Schuurmann G. Multivariate Discrimination between Modes of Toxic Action of Phenols // Quant. Struct.-Act. Relat. 2002. Vol. 21. № 1. P. 12-22.

36. Zhao Y.H., Zhang X.J., Wen Y., Sun F.T., Guo Z., Qin W.C., Qin H.W., Xu J.L., Sheng L.X., Abraham M.H. Toxicity of Organic Chemicals to *Tetrahymena pyriformis*: Effect of Polarity and Ionization on Toxicity // *Chemosphere*. 2010. Vol. 79. № 1. P. 72-77.
37. Burden F.R., Winkler D.A. A Quantitative Structure-Activity Relationships Model for the Acute Toxicity of Substituted Benzenes to *Tetrahymena pyriformis* Using Bayesian-Regularized Neural Networks // *Chem. Res. Toxicol.* 2000. Vol. 13. № 6. P. 436-440.
38. Burden F.R. Quantitative Structure-Activity Relationship Studies Using Gaussian Processes // *J. Chem. Inf. Comput. Sci.* 2001. Vol. 41. № 3. P. 830-835.
39. Netzeva T.I., Schultz T.W., Aptula A.O., Cronin M.T.D. Partial Least Squares Modelling of the Acute Toxicity of Aliphatic Compounds to *Tetrahymena pyriformis* // *SAR QSAR Environ. Res.* 2003. Vol. 14. № 4. P. 265-283.
40. Estrada E., Uriarte E. Quantitative Structure-Toxicity Relationships Using Tops-Mode. 1. Nitrobenzene Toxicity to *Tetrahymena pyriformis* // *SAR QSAR Environ. Res.* 2001. Vol. 12. № 3. P. 309-324.
41. Xue Y., Li H., Ung C.Y., Yap C.W., Chen Y.Z. Classification of a Diverse Set of *Tetrahymena pyriformis* Toxicity Chemical Compounds from Molecular Descriptors by Statistical Learning Methods // *Chem. Res. Toxicol.* 2006. Vol. 19. № 8. P. 1030-1039.
42. Kahn I., Sild S., Maran U. Modeling the Toxicity of Chemicals to *Tetrahymena pyriformis* Using Heuristic Multilinear Regression and Heuristic Back-Propagation Neural Networks // *J. Chem. Inf. Model.* 2007. Vol. 47. № 6. P. 2271-2279.
43. Duchowicz P.R., Mercader A.G., Fernández F.M., Castro E.A. Prediction of Aqueous Toxicity for Heterogeneous Phenol Derivatives by QSAR // *Chemometr. Intell. Lab. Syst.* 2008. Vol. 90. № 2. P. 97-107.

44. Yao X.J., Panaye A., Doucet J.P., Chen H.F., Zhang R.S., Fan B.T., Liu M.C., Hu Z.D. Comparative Classification Study of Toxicity Mechanisms Using Support Vector Machines and Radial Basis Function Neural Networks // *Analyt. Chim. Acta* 2005. Vol. 535. № 1-2. P. 259-273.
45. Toropov A.A., Schultz T.W. Prediction of Aquatic Toxicity: Use of Optimization of Correlation Weights of Local Graph Invariants // *J. Chem. Inf. Comput. Sci.* 2003. Vol. 43. № 2. P. 560-567.
46. Roy D.R., Parthasarathi R., Subramanian V., Chattaraj P.K. An Electrophilicity Based Analysis of Toxicity of Aromatic Compounds Towards *Tetrahymena pyriformis* // *QSAR Comb. Sci.* 2006. Vol. 25. № 2. P. 114-122.
47. Mohajeri A., Dinpajoo M.H. Structure–Toxicity Relationship for Aliphatic Compounds Using Quantum Topological Descriptors // *J. Mol. Struct. (Theochem)*. 2008. Vol. 855. № 1-3. P. 1-5.
48. Kar S., Harding A.P., Roy K., Popelier P.L.A. QSAR with Quantum Topological Molecular Similarity Indices: Toxicity of Aromatic Aldehydes to *Tetrahymena pyriformis* // *SAR QSAR Environ. Res.* 2010. Vol. 21. № 1-2. P. 149-168.
49. Katritzky A.R., Oliferenko P., Oliferenko A., Lomaka A., Karelson M. Nitrobenzene Toxicity: QSAR Correlations and Mechanistic Interpretations // *J. Phys. Org. Chem.* 2003. Vol. 16. № 10. P. 811-817.
50. Spycher S., Pellegrini E., Gasteiger J. Use of Structure Descriptors To Discriminate between Modes of Toxic Action of Phenols // *J. Chem. Inf. Model.* 2005. Vol. 45. № 1. P. 200-208.
51. Tetko I.V., Sushko I., Pandey A.K., Zhu H., Tropsha A., Papa E., Oberg T., Todeschini R., Fourches D., Varnek A. Critical Assessment of QSAR Models of Environmental Toxicity against *Tetrahymena pyriformis*: Focusing on Applicability Domain and Overfitting by Variable Selection // *J. Chem. Inf. Model.* 2008. Vol. 48. № 9. P. 1733-1746.
52. Kahn I., Maran U., Benfenati E., Netzeva T.I., Schultz T.W., Cronin M.T.D. Comparative Quantitative Structure–Activity–Activity Relationships for

- Toxicity to *Tetrahymena pyriformis* and *Pimephales promelas* // ATLA. 2007. Vol. 35. № 1. P. 15-24.
53. Zhao Y.H., Yuan X., Su L.M., Qin W.C., Abraham M.H. Classification of Toxicity of Phenols to *Tetrahymena pyriformis* and Subsequent Derivation of QSARs from Hydrophobic, Ionization and Electronic Parameters // Chemosphere. 2009. Vol. 75. № 7. P. 866-871.
54. Dimitrov S.D., Mekenyan O.G., Schultz T.W. Interspecies Modeling of Narcotics Toxicity to Aquatic Animals // Bull. Environ. Contam. Toxicol. 2000. Vol. 65. № 3. P. 399-406.
55. Dimitrov S., Koleva Y., Schultz T.W., Walker J.D., Mekenyan O. Interspecies Quantitative Structure–Activity Relationship Model for Aldehydes: Aquatic Toxicity // Environ. Toxicol. Chem. 2004. Vol. 23. № 2. P. 463-470.
56. Ivanciuc O. Support Vector Machines Prediction of the Mechanism of Toxic Action from Hydrophobicity and Experimental Toxicity Against *Pimephales promelas* and *Tetrahymena pyriformis* // Internet Electron. J. Mol. Des. 2004. Vol. 3. № 12. P. 802–821.
57. Netzeva T.I., Schultz T.W. QSARs for the Aquatic Toxicity of Aromatic Aldehydes from *Tetrahymena* Data // Chemosphere. 2005. Vol. 61. № 11. P. 1632-1643.
58. Roy D.R., Parthasarathi R., Maiti B., Subramanian V., Chattaraj P.K. Electrophilicity as a Possible Descriptor for Toxicity Prediction // Bioorg. Med. Chem. 2005. Vol. 13. № 10. P. 3405-3412.
59. Toropov A.A., Benfenati E. QSAR Modelling of Aldehyde Toxicity Against a Protozoan, *Tetrahymena pyriformis* by Optimization of Correlation Weights of Nearest Neighboring Codes // J. Mol. Struct. (Theochem). 2004. Vol. 679. № 3. P. 225-228.
60. DeWeese A.D., Schultz T.W. Structure–Activity Relationships for Aquatic Toxicity to *Tetrahymena*: Halogen-Substituted Aliphatic Esters // Environ. Toxicol. 2001. Vol. 16. № 1. P. 54-60.

61. Li X., Zhang T., Min X., Liu P. Toxicity of Aromatic Compounds to Tetrahymena Estimated by Microcalorimetry and QSAR // *Aquat. Toxicol.* 2010. Vol. 98. № 4. P. 322-327.
62. Dimitrov S.D., Mekenyan O.G., Sinks G.D., Schultz T.W. Global Modeling of Narcotic Chemicals: Ciliate and Fish Toxicity // *J. Mol. Struct. (Theochem)*. 2003. Vol. 622. № 1-2. P. 63-70.
63. Schultz T.W., Yarbrough J.W., Pilkington T.B. Aquatic Toxicity and Abiotic Thiol Reactivity of Aliphatic Isothiocyanates: Effects of Alkyl-Size and -Shape // *Environ. Toxicol. Phar.* 2007. Vol. 23. № 1. P. 10-17.
64. Aptula A.O., Roberts D.W., Cronin M.T.D., Schultz T.W. Chemistry-Toxicity Relationships for the Effects of Di- and Trihydroxybenzenes to Tetrahymena pyriformis // *Chem. Res. Toxicol.* 2005. Vol. 18. № 5. P. 844-854.
65. Ren S. Prediction of the Mechanisms of Toxic Action of Phenols: Baseline-Category Logit Models // *Aquat. Toxicol.* 2003. Vol. 65. № 1. P. 111-115.
66. Castillo-Garit J.A., Marrero-Ponce Y., Escobar J., Torrens F., Rotondo R. A Novel Approach to Predict Aquatic Toxicity from Molecular Structure // *Chemosphere*. 2008. Vol. 73. № 3. P. 415-427.
67. Ren S., Kim H. Comparative Assessment of Multiresponse Regression Methods for Predicting the Mechanisms of Toxic Action of Phenols // *J. Chem. Inf. Comput. Sci.* 2003. Vol. 43. № 6. P. 2106-2110.
68. Ren S. Phenol Mechanism of Toxic Action Classification and Prediction: a Decision Tree Approach // *Toxicol. Lett.* 2003. Vol. 144. № 3. P. 313-323.
69. Agrawal V.K., Khadikar P.V. QSAR Prediction of Toxicity of Nitrobenzenes // *Bioorg. Med. Chem.* 2001. Vol. 9. № 11. P. 3035-3040.
70. Ren S. Ecotoxicity Prediction Using Mechanism- and Non-Mechanism-Based QSARs: a Preliminary Study // *Chemosphere*. 2003. Vol. 53. № 9. P. 1053-1065.
71. Devillers J. Linear Versus Nonlinear QSAR Modeling of the Toxicity of Phenol Derivatives to Tetrahymena pyriformis // *SAR QSAR Environ. Res.* 2004. Vol. 15. № 4. P. 237-249.

72. Schuurmann G., Aptula A.O., Kuhne R., Ebert R.U. Stepwise Discrimination between Four Modes of Toxic Action of Phenols in the *Tetrahymena pyriformis* Assay // Chem. Res. Toxicol. 2003. Vol. 16. № 8. P. 974-987.
73. Enoch S.J., Cronin M.T.D., Schultz T.W., Madden J.C. An Evaluation of Global QSAR Models for the Prediction of the Toxicity of Phenols to *Tetrahymena pyriformis* // Chemosphere. 2008. Vol. 71. № 7. P. 1225-1232.
74. Jalali-Heravi M., Kyani A. Comparative Structure–Toxicity Relationship Study of Substituted Benzenes to *Tetrahymena pyriformis* Using Shuffling-Adaptive Neuro Fuzzy Inference System and Artificial Neural Networks // Chemosphere. 2008. Vol. 72. № 5. P. 733-740.
75. Roy K., Ghosh G. QSTR with Extended Topochemical Atom (ETA) Indices. 12. QSAR for the Toxicity of Diverse Aromatic Compounds to *Tetrahymena pyriformis* Using Chemometric Tools // Chemosphere. 2009. Vol. 77. № 7. P. 999-1009.
76. Roy K., Das R.N. QSTR with Extended Topochemical Atom (ETA) Indices. 14. QSAR Modeling of Toxicity of Aromatic Aldehydes to *Tetrahymena pyriformis* // J. Hazard. Mater. 2010. Vol. 183. № 1-3. P. 913-922.
77. Cheng F., Shen J., Yu Y., Li W., Liu G., Lee P.W., Tang Y. In Silico Prediction of *Tetrahymena pyriformis* Toxicity for Diverse Industrial Chemicals with Substructure Pattern Recognition and Machine Learning Methods // Chemosphere. 2011. Vol. 82. № 11. P. 1636-1643.
78. Roy K., Ghosh G. QSTR with Extended Topochemical Atom Indices. 3. Toxicity of Nitrobenzenes to *Tetrahymena pyriformis* // QSAR Comb. Sci. 2004. Vol. 23. № 2-3. P. 99-108.
79. Zhang X.J., Qin H.W., Su L.M., Qin W.C., Zou M.Y., Sheng L.X., Zhao Y.H., Abraham M.H. Interspecies Correlations of Toxicity to Eight Aquatic Organisms: Theoretical Considerations // Sci. Total Environ. 2010. Vol. 408. № 20. P. 4549-4555.

80. Zvinavashe E., Murk A.J., Vervoort J., Soffers A.E.M.F., Freidig A., Rietjens I.M.C.M. Quantum Chemistry Based Quantitative Structure-Activity Relationships for Modeling the (Sub)Acute Toxicity of Substituted Mononitrobenzenes in Aquatic Systems // *Environ. Toxicol. Chem.* 2006. Vol. 25. № 9. P. 2313-2321.
81. Boeije G.M., Cano M.L., Marshall S.J., Belanger S.E., Van Compernelle R., Dorn P.B., Gumbel H., Toy R., Wind T. Ecotoxicity Quantitative Structure-Activity Relationships for Alcohol Ethoxylate Mixtures Based on Substance-Specific Toxicity Predictions // *Ecotoxicol. Environ. Saf.* 2006. Vol. 64. № 1. P. 75-84.
82. Basak S.C., Grunwald G.D., Gute B.D., Balasubramanian K., Opitz D. Use of Statistical and Neural Net Approaches in Predicting Toxicity of Chemicals // *J. Chem. Inf. Comput. Sci.* 2000. Vol. 40. № 4. P. 885-890.
83. Kaiser K.L.E., Niculescu S.P. Using Probabilistic Neural Networks to Model the Toxicity of Chemicals to the Fathead minnow (*Pimephales promelas*): A Study Based on 865 Compounds // *Chemosphere.* 1999. Vol. 38. № 14. P. 3237-3245.
84. Lozano S., Lescot E., Halm M.P., Lepailleur A., Bureau R., Rault S. Prediction of Acute Toxicity in Fish by Using QSAR Methods and Chemical Modes of Action // *J. Enzyme Inhib. Med. Chem.* 2010. Vol. 25. № 2. P. 195-203.
85. Klopman G., Saiakhov R., Rosenkranz H.S. Multiple Computer-Automated Structure Evaluation Study of Aquatic Toxicity II. Fathead minnow // *Environ. Toxicol. Chem.* 2000. Vol. 19. № 2. P. 441-447.
86. Di Marzio W., Galassi S., Todeschini R., Consolaro F. Traditional versus WHIM Molecular Descriptors in QSAR Approaches Applied to Fish Toxicity Studies // *Chemosphere.* 2001. Vol. 44. № 3. P. 401-406.
87. Kramer N.I., Hermens J.L.M., Schirmer K. The Influence of Modes of Action and Physicochemical Properties of Chemicals on the Correlation Between in Vitro and Acute Fish Toxicity Data // *Toxicology in Vitro.* 2009. Vol. 23. № 7. P. 1372-1379.

88. Yan X.F., Xiao H.M., Gong X.D., Ju X.H. A Comparison of Semiempirical and First Principle Methods for Establishing Toxicological QSARs of Nitroaromatics // *J. Mol. Struct. (Theochem)*. 2006. Vol. 764. № 1-3. P. 141-148.
89. Huuskonen J. QSAR Modeling with the Electrotopological State Indices: Predicting the Toxicity of Organic Chemicals // *Chemosphere*. 2003. Vol. 50. № 7. P. 949-953.
90. Michielan L., Pireddu L., Floris M., Moro S. Support Vector Machine (SVM) as Alternative Tool to Assign Acute Aquatic Toxicity Warning Labels to Chemicals // *Mol. Inf.* 2010. Vol. 29. № 1-2. P. 51-64.
91. Lozano S., Halm-Lemeille M.P., Lepailleur A., Rault S., Bureau R. Consensus QSAR Related to Global or MOA Models: Application to Acute Toxicity for Fish // *Mol. Inf.* 2010. Vol. 29. № 11. P. 803-813.
92. Basak S.C., Gute B.D., Lucic B., Nikolic S., Trinajstić N. A Comparative QSAR Study of Benzamidines Complement-Inhibitory Activity and Benzene Derivatives Acute Toxicity // *Comput. Chem.* 2000. Vol. 24. № 2. P. 181-191.
93. Öberg T. A QSAR for Baseline Toxicity: Validation, Domain of Application, and Prediction // *Chem. Res. Toxicol.* 2004. Vol. 17. № 12. P. 1630-1637.
94. Colombo A., Benfenati E., Karelson M., Maran U. The Proposal of Architecture for Chemical Splitting to Optimize QSAR Models for Aquatic Toxicity // *Chemosphere*. 2008. Vol. 72. № 5. P. 772-780.
95. Yuan H., Wang Y.Y., Cheng Y.Y. Mode of Action-Based Local QSAR Modeling for the Prediction of Acute Toxicity in the Fathead minnow // *J. Mol. Graph. Model.* 2007. Vol. 26. № 1. P. 327-335.
96. Yuan H., Wang Y., Cheng Y. Local and Global Quantitative Structure-Activity Relationship Modeling and Prediction for the Baseline Toxicity // *J. Chem. Inf. Model.* 2007. Vol. 47. № 1. P. 159-169.
97. Papa E., Villa F., Gramatica P. Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in

- Pimephales promelas (Fathead Minnow) // J. Chem. Inf. Model. 2005. Vol. 45. № 5. P. 1256-1266.
98. Tan N.X., Li P., Rao H.B., Li Z.R., Li X.Y. Prediction of the Acute Toxicity of Chemical Compounds to the Fathead minnow by Machine Learning Approaches // Chemometr. Intell. Lab. Syst. 2010. Vol. 100. № 1. P. 66-73.
99. Netzeva T.I., Aptula A.O., Benfenati E., Cronin M.T.D., Gini G., Lessigiarska I., Maran U., Vracko M., Schuurmann G. Description of the Electronic Structure of Organic Chemicals Using Semiempirical and Ab Initio Methods for Development of Toxicological QSARs // J. Chem. Inf. Model. 2005. Vol. 45. № 1. P. 106-114.
100. Pintore M., Piclin N., Benfenati E., Gini G., Chretien J.R. Predicting Toxicity Against the Fathead minnow by Adaptive Fuzzy Partition // QSAR Comb. Sci. 2003. Vol. 22. № 2. P. 210-219.
101. Smiesko M., Benfenati E. Predictive Models for Aquatic Toxicity of Aldehydes Designed for Various Model Chemistries // J. Chem. Inf. Comput. Sci. 2004. Vol. 44. № 3. P. 976-984.
102. Lodhi H., Muggleton S., Sternberg M.J.E. Multi-class Mode of Action Classification of Toxic Compounds Using Logic Based Kernel Methods // Mol. Inf. 2010. Vol. 29. № 8-9. P. 655-664.
103. Niculescu S.P., Atkinson A., Hammond G., Lewis M. Using Fragment Chemistry Data Mining and Probabilistic Neural Networks in Screening Chemicals for Acute Toxicity to the Fathead minnow // SAR QSAR Environ. Res. 2004. Vol. 15. № 4. P. 293-309.
104. Toropov A.A., Toropova A.P. QSAR Modeling of Toxicity on Optimization of Correlation Weights of Morgan Extended Connectivity // J. Mol. Struct. (Theochem). 2002. Vol. 578. № 1-3. P. 129-134.
105. Mazzatorta P., Benfenati E., Neagu C.D., Gini G. Tuning Neural and Fuzzy-Neural Networks for Toxicity Modeling // J. Chem. Inf. Comput. Sci. 2003. Vol. 43. № 2. P. 513-518.

106. Moore D.R.J., Breton R.L., Macdonald D.B. A Comparison of Model Performance for Six Quantitative Structure-Activity Relationship Packages That Predict Acute Toxicity to Fish // *Environ. Toxicol. Chem.* 2003. Vol. 22. № 8. P. 1799-1809.
107. Dyer S.D., Versteeg D.J., Belanger S.E., Chaney J.G., Mayer F.L. Interspecies Correlation Estimates Predict Protective Environmental Concentrations // *Environ. Sci. Technol.* 2006. Vol. 40. № 9. P. 3102-3111.
108. Freidig A.P., Hermens J.L.M. Narcosis and Chemical Reactivity QSARs for Acute Fish Toxicity // *Quant. Struct.-Act. Relat.* 2000. Vol. 19. № 6. P. 547-553.
109. Faucon J.C., Bureau R., Faisant J., Briens F., Rault S. Prediction of the Fish Acute Toxicity from Heterogeneous Data Coming from Notification Files // *Chemosphere.* 1999. Vol. 38. № 14. P. 3261-3276.
110. Lammer E., Carr G.J., Wendler K., Rawlings J.M., Belanger S.E., Braunbeck Th. Is the Fish Embryo Toxicity Test (FET) with the Zebrafish (*Danio rerio*) a Potential Alternative for the Fish Acute Toxicity Test? // *Comp. Biochem. Physiol. Part C.* 2009. Vol. 149. № 2. P. 196-209.
111. Khadikar P.V., Mather K.C., Singh S., Phadnis A., Shrivastava A., Mandaloi M. Study on Quantitative Structure-Toxicity Relationships of Benzene Derivatives Acting by Narcosis // *Bioorg. Med. Chem.* 2002. Vol. 10. № 6. P. 1761-1766.
112. Klopman G., Saiakhov R., Rosenkranz H.S., Hermens J.L.M. Multiple Computer-Automated Structure Evaluation Program Study of Aquatic Toxicity 1: Guppy // *Environ. Toxicol. Chem.* 1999. Vol. 18. № 11. P. 2497-2505.
113. Rose K., Hall L.H. E-State Modeling of Fish Toxicity Independent of 3D Structure Information // *SAR QSAR Environ. Res.* 2003. Vol. 14. № 2. P. 113-129.
114. Roy K., Ghosh G. QSTR with Extended Topochemical Atom Indices. 2. Fish Toxicity of Substituted Benzenes // *J. Chem. Inf. Comput. Sci.* 2004. Vol. 44. № 2. P. 559-567.

115. Zvinavashe E., van den Berg H., Soffers A.E.M.F., Vervoort J., Freidig A., Murk A.J., Rietjens I.M.C.M. QSAR Models for Predicting in Vivo Aquatic Toxicity of Chlorinated Alkanes to Fish // *Chem. Res. Toxicol.* 2008. Vol. 21. № 3. P. 739-745.
116. Knauer K., Lampert C., Gonzalez-Valero J. Comparison of in Vitro and in Vivo Acute Fish Toxicity in Relation to Toxicant Mode of Action // *Chemosphere.* 2007. Vol. 68. № 8. P. 1435-1441.
117. Mazzatorta P., Smiesko M., Piparo E.L., Benfenati E. QSAR Model for Predicting Pesticide Aquatic Toxicity // *J. Chem. Inf. Model.* 2005. Vol. 45. № 6. P. 1767-1774.
118. Bermudez-Saldana J.M., Escuder-Gilabert L., Medina-Hernandez M.J., Villanueva-Camanas R.M., Sagrado S. Chromatographic Evaluation of the Toxicity in Fish of Pesticides // *J. Chromatogr. B.* 2005. Vol. 814. № 1. P. 115-125.
119. Tao S., Xi X., Fuliu Xu, Dawson R. A QSAR Model for Predicting Toxicity (LC_{50}) to Rainbow trout // *Water Res.* 2002. Vol. 36. № 11. P. 2926-2930.
120. Toropov A.A., Benfenati E. Correlation Weighting of Valence Shells in QSAR Analysis of Toxicity // *Bioorg. Med. Chem.* 2006. Vol. 14. № 11. P. 3923-3928.
121. Bermudez-Saldana J.M., Cronin M.T.D. Quantitative Structure–Activity Relationships for the Toxicity of Organophosphorus and Carbamate Pesticides to the Rainbow trout *Onchorhynchus mykiss* // *Pest. Manag. Sci.* 2006. Vol. 62. № 9. P. 819-831.
122. Devogelaere D., Van Bael P., Rijckaert M. A Soft Computing Approach for Toxicity Prediction // *Lecture Notes in Computer Science.* 2000. Vol. 1952. P. 437-446.
123. Delistraty D. Acute Toxicity to Rats and Trout with a Focus on Inhalation and Aquatic Exposures // *Ecotoxicol. Environ. Saf.* 2000. Vol. 46. № 2. P. 225-233.

124. Casalegno M., Sello G., Benfenati E. Top-Priority Fragment QSAR Approach in Predicting Pesticide Aquatic Toxicity // Chem. Res. Toxicol. 2006. Vol. 19. № 11. P. 1533-1539.
125. Zhao C.Y., Zhang H.X., Zhang X.Y., Liu M.C., Hu Z.D., Fan B.T. Application of Support Vector Machine (SVM) for Prediction Toxic Activity of Different Data Sets // Toxicol. 2006. Vol. 217. № 2-3. P. 105-119.
126. Xu S., Li L., Tan Y., Feng J., Wei Z., Wang L. Prediction and QSAR Analysis of Toxicity to Photobacterium phosphoreum for a Group of Heterocyclic Nitrogen Compounds // Bull. Environ. Contam. Toxicol. 2000. Vol. 64. № 3. P. 316-322.
127. Lin Z., Yu H., Wei D., Wang G., Feng J., Wang L. Prediction of Mixture Toxicity with its Total Hydrophobicity // Chemosphere. 2002. Vol. 46. № 2. P. 305-310.
128. Lin Z., Zhong P., Yin K., Wang L., Yu H. Quantification of Joint Effect for Hydrogen Bond and Development of QSARs for Predicting Mixture Toxicity // Chemosphere. 2003. Vol. 52. № 7. P. 1199-1208.
129. Liu X., Wu C., Han S., Wang L., Zhang Z. The Acute Toxicity of α -Branched Phenylsulfonyl Acetates in Photobacterium phosphoreum Test // Ecotoxicol. Environ. Saf. 2001. Vol. 49. № 3. P. 240-244.
130. Lin Z., Yin K., Shi P., Wang L., Yu H. Development of QSARs for Predicting the Joint Effects between Cyanogenic Toxicants and Aldehydes // Chem. Res. Toxicol. 2003. Vol. 16. № 10. P. 1365-1371.
131. Liu X., Yang Z., Wang L. Three-Dimensional Quantitative Structure–Activity Relationship Study for Phenylsulfonyl Carboxylates Using CoMFA and CoMSIA // Chemosphere. 2003. Vol. 53. № 8. P. 945-952.
132. Wei D.B., Zhai L.H., Hu H.-Y. Qsar-Based Toxicity Classification and Prediction for Single and Mixed Aromatic Compounds // SAR QSAR Environ. Res. 2004. Vol. 15. № 3. P. 207-216.

133. Chen Q., Kou Y.W., Wang Q., Chen H., Yuan J. A Molecular Fragments Variable Connectivity Index for Studying the Toxicity (*Vibrio fischeri* pT₅₀) of Substituted-Benzenes // J. Environ. Sci. Health Part A. 2009. Vol. 44. № 3. P. 288-294.
134. Agrawal V.K., Khadikar P.V. QSAR Study on Narcotic Mechanism of Action and Toxicity: a Molecular Connectivity Approach to *Vibrio fischeri* Toxicity Testing // Bioorg. Med. Chem. 2002. Vol. 10. № 11. P. 3517-3522.
135. Parvez S., Venkataraman C., Mukherji S. Toxicity Assessment of Organic Pollutants: Reliability of Bioluminescence Inhibition Assay and Univariate QSAR Models Using Freshly Prepared *Vibrio fischeri* // Toxicology in Vitro. 2008. Vol. 22. № 7. P. 1806-1813.
136. Cronin M.T.D., Bowers G.S., Sinks G.D., Schultz T.W. Structure-Toxicity Relationships for Aliphatic Compounds Encompassing a Variety of Mechanisms of Toxic Action to *Vibrio fischeri* // SAR QSAR Environ. Res. 2000. Vol. 11. № 3. P. 301-312.
137. Lu G.H., Yuan X., Wang C. Quantitative Structure–Toxicity Relationships for Substituted Aromatic Compounds to *Vibrio fischeri* // Bull. Environ. Contam. Toxicol. 2003. Vol. 70. № 4. P. 832-838.
138. Roy K., Ghosh G. QSTR with Extended Topochemical Atom Indices. Part 5: Modeling of the Acute Toxicity of Phenylsulfonyl Carboxylates to *Vibrio fischeri* Using Genetic Function Approximation // Bioorg. Med. Chem. 2005. Vol. 13. № 4. P. 1185-1194.
139. Tchounwou P.B., Wilson B., Ishaque A., Ransome R., Huang M.-J., Leszczynski J. Toxicity Assessment of Atrazine and Related Triazine Compounds in the Microtox Assay, and Computational Modeling for Their Structure-Activity Relationship // Int. J. Mol. Sci. 2000. Vol. 1. № 4. P. 63-74.
140. Warne M.A., Osborn D., Lindon J.C., Nicholson J.K. Quantitative Structure-Toxicity Relationships for Halogenated Substituted-Benzenes to *Vibrio fischeri*,

- Using Atom-Based Semi-Empirical Molecular-Orbital Descriptors // *Chemosphere*. 1999. Vol. 38. № 14. P. 3357-3382.
141. Roy K., Ghosh G. QSTR with Extended Topochemical Atom Indices. 4. Modeling of the Acute Toxicity of Phenylsulfonyl Carboxylates to *Vibrio fischeri* using Principal Component Factor Analysis and Principal Component Regression Analysis // *QSAR Comb. Sci.* 2004. Vol. 23. № 7. P. 526-535.
142. Liu X., Yang Z., Wang L. COMFA of the Acute Toxicity of Phenylsulfonyl Carboxylates to *Vibrio fischeri* // *SAR QSAR Environ. Res.* 2003. Vol. 14. № 3. P. 183-190.
143. Cronin M.T.D., Zhao Y.H., Yu R.L. pH-Dependence and QSAR Analysis of the Toxicity of Phenols and Anilines to *Daphnia magna* // *Environ. Toxicol.* 2000. Vol. 15. № 2. P. 140-148.
144. Von der Ohe P.C., Kuhne R., Ebert R.W., Altenburger R., Liess M., Schuurmann G. Structural Alerts - A New Classification Model to Discriminate Excess Toxicity from Narcotic Effect Levels of Organic Compounds in the Acute Daphnid Assay // *Chem. Res. Toxicol.* 2005. Vol. 18. № 3. P. 536-555.
145. Liu X., Wang B., Huang Z., Han S., Wang L. Acute Toxicity and Quantitative Structure–Activity Relationships of α -Branched Phenylsulfonyl Acetates to *Daphnia magna* // *Chemosphere*. 2003. Vol. 50. № 3. P. 403-408.
146. Tao S., Xi X., Xu F., Li B., Cao J., Dawson R. A Fragment Constant QSAR Model for Evaluating the EC₅₀ Values of Organic Chemicals to *Daphnia magna* // *Environ. Pollut.* 2002. Vol. 116. № 1. P. 57-64.
147. Dai J., Jin L., Yao S., Wang L. Prediction of Partition Coefficient and Toxicity for Benzaldehyde Compounds by their Capacity Factors and Various Molecular Descriptors // *Chemosphere*. 2001. Vol. 42. № 8. P. 899-907.
148. Faucon J.C., Bureau R., Faisant J., Briens F., Rault S. Prediction of the *Daphnia* Acute Toxicity from Heterogeneous Data // *Chemosphere*. 2001. Vol. 44. № 3. P. 407-422.

149. Kar S., Roy K. QSAR Modeling of Toxicity of Diverse Organic Chemicals to *Daphnia magna* Using 2D and 3D Descriptors // *J. Hazard. Mater.* 2010. Vol. 177. № 1-3. P. 344-351.
150. Toropov A.A., Benfenati E. QSAR Models for *Daphnia* Toxicity of Pesticides Based on Combinations of Topological Parameters of Molecular Structures // *Bioorg. Med. Chem.* 2006. Vol. 14. № 8. P. 2779-2788.
151. Papa E., Battaini F., Gramatica P. Ranking of Aquatic Toxicity of Esters Modelled by QSAR // *Chemosphere.* 2005. Vol. 58. № 5. P. 559-570.
152. Kar S., Roy K. First Report on Interspecies Quantitative Correlation of Ecotoxicity of Pharmaceuticals // *Chemosphere.* 2010. Vol. 81. № 6. P. 738-747.
153. De Roode D., Hoekzema C., De Vries-Buitenweg S., Van de Waart B., Van der Hoeven J. QSARs in Ecotoxicological Risk Assessment // *Regul. Toxicol. Pharmacol.* 2006. Vol. 45. № 1. P. 24-35.
154. Sanderson H., Thomsen M. Comparative Analysis of Pharmaceuticals versus Industrial Chemicals Acute Aquatic Toxicity Classification According to the United Nations Classification System for Chemicals // *Toxicol. Lett.* 2009. Vol. 187. № 2. P. 84-93.
155. Fisk P.R., Wildey R.J., Girling A.E., Sanderson H., Belanger S.E., Veenstra G., Nielsen A., Kasai Y., Willing A., Dyer S.D., Stanton K. Environmental Properties of Long Chain Alcohols. Part 1: Physicochemical, Environmental Fate and Acute Aquatic Toxicity Properties // *Ecotoxicol. Environ. Saf.* 2009. Vol. 72. № 4. P. 980-995.
156. Cui S., Wang X., Liu S., Wang L. Predicting Toxicity of Benzene Derivatives by Molecular Hologram Derived Quantitative Structure-Activity Relationships (QSARS) // *SAR QSAR Environ. Res.* 2003. Vol. 14. № 3. P. 223-231.
157. Netzeva T.I., Dearden J.C., Edwards R., Worgan A.D.P., Cronin M.T.D. QSAR Analysis of the Toxicity of Aromatic Compounds to *Chlorella Vulgaris* in a Novel Short-Term Assay // *J. Chem. Inf. Comput. Sci.* 2004. Vol. 44. № 1. P. 258-265.

158. Shi W., Zhang X., Shen Q. Quantitative Structure-Activity Relationships Studies of CCR5 Inhibitors and Toxicity of Aromatic Compounds Using Gene Expression Programming // *Eur. J. Med. Chem.* 2010. Vol. 45. № 1. P. 49-54.
159. Shen Q., Shi W.M., Kong W. Modified Tabu Search Approach for Variable Selection in Quantitative Structure-Activity Relationship Studies of Toxicity of Aromatic Compounds // *Artif. Intell. Med.* 2010. Vol. 49. № 1. P. 61-66.
160. Roy K., Ghosh G. QSTR with Extended Topochemical Atom (ETA) Indices. 9. Comparative QSAR for the Toxicity of Diverse Functional Organic Compounds to *Chlorella vulgaris* Using Chemometric Tools // *Chemosphere.* 2007. Vol. 70. № 1. P. 1-12.
161. Xia B., Liu K., Gong Z., Zheng B., Zhang X., Fan B. Rapid Toxicity Prediction of Organic Chemicals to *Chlorella vulgaris* Using Quantitative Structure-Activity Relationships Methods // *Ecotoxicol. Environ. Saf.* 2009. Vol. 72. № 3. P. 787-794.
162. Guida X., Jianhua H., Xiaomin L. Synthesis and QSAR Studies of Novel 1-substituted-2-aminobenzimidazoles Derivatives // *Eur. J. Med. Chem.* 2006. Vol. 41. № 9. P. 1080-1083.
163. Randić M., Basak S.C. On Use of the Variable Connectivity Index χ in QSAR: Toxicity of Aliphatic Ethers // *J. Chem. Inf. Comput. Sci.* 2001. Vol. 41. № 3. P. 614-618.
164. Li Z., Sun Y., Yan X., Meng F. Study on QSTR of Benzoic Acid Compounds with MCI // *Int. J. Mol. Sci.* 2010. Vol. 11. № 4. P. 1228-1235.
165. Papa E., Luini M., Gramatica P. Quantitative Structure-Activity Relationship Modelling of Oral Acute Toxicity and Cytotoxic Activity of Fragrance Materials in Rodents // *SAR QSAR Environ. Res.* 2009. Vol. 20. № 7-8. P. 767-779.
166. Sedykh A., Zhu H., Tang H., Zhang L., Richard A., Rusyn I., Tropsha A. Use of in Vitro HTS-Derived Concentration-Response Data as Biological Descriptors Improves the Accuracy of QSAR Models of in Vivo Toxicity // *Environ. Health Perspect.* 2011. Vol. 119. № 3. P. 364-370.

167. Carlsen L., Kenessov B.N., Batyrbekova S.Y. A QSAR/QSTR Study on the Human Health Impact of the Rocket Fuel 1,1-dimethyl hydrazine and its Transformation Products: Multicriteria Hazard Ranking Based on Partial Order Methodologies // *Environ. Toxicol. Phar.* 2009. Vol. 27. № 3. P. 415-423.
168. Guilian W., Naibin B. Structure-Activity Relationships for Rat and Mouse LD50 of Miscellaneous Alcohols // *Chemosphere*. 1998. Vol. 36. № 7. P. 1475-1483.
169. Hau K.M., Connell D.W., Richardson B.J. Mechanism of Acute Inhalation Toxicity of Alkanes and Aliphatic Alcohols // *Environ. Toxicol. Phar.* 1999. Vol. 7. № 3. P. 159-167.
170. Koleva Y.K., Cronin M.T.D., Madden J.C., Schwöbel J.A.H. Modelling Acute Oral Mammalian Toxicity. 1. Definition of a Quantifiable Baseline Effect // *Toxicology in Vitro*. 2011. Vol. 25. № 7. P. 1281-1293.
171. Veith G.D., Petkova E.P., Wallace K.B. A Baseline Inhalation Toxicity Model for Narcosis in Mammals // *SAR QSAR Environ. Res.* 2009. Vol. 20. № 5-6. P. 567-578.
172. Sazonovas A., Japertas P., Didziapetris R. Estimation of Reliability of Predictions and Model Applicability Domain Evaluation in the Analysis of Acute Toxicity (LD₅₀) // *SAR QSAR Environ. Res.* 2010. Vol. 21. № 1-2. P. 127-148.
173. Zhu H., Ye L., Richard A., Golbraikh A., Wright F.A., Rusyn I., Tropsha A. A Novel Two-Step Hierarchical Quantitative Structure–Activity Relationship Modeling Work Flow for Predicting Acute Toxicity of Chemicals in Rodents // *Environ. Health Perspect.* 2009. Vol. 117. № 8. P. 1257-1264.
174. Jean P.A., Gallavan R.H., Kolesar G.B., Siddiqui W.H., Oxley J.A., Meeks R.G. Chlorosilane Acute Inhalation Toxicity and Development of an LC50 Prediction Model // *Inhal. Toxicol.* 2006. Vol. 18. № 8. P. 515-522.
175. Rasulev B., Kušić H., Leszczynska D., Leszczynski J., Koprivanac N. QSAR Modeling of Acute Toxicity on Mammals Caused by Aromatic Compounds: the

- Case Study Using Oral LD₅₀ for Rats // J. Environ. Monit. 2010. Vol. 12. № 5. P. 1037-1044.
176. Juranić I.O., Drakulić B.J., Petrović S.D., Mijin D.Z., Stanković M.V. A QSAR Study of Acute Toxicity of N-Substituted Fluoroacetamides to Rats // Chemosphere. 2006. Vol. 62. № 4. P. 641-649.
177. Toropov A.A., Rasulev B.F., Leszczynski J. QSAR Modeling of Acute Toxicity for Nitrobenzene Derivatives Towards Rats: Comparative Analysis by MLRA and Optimal Descriptors // QSAR Comb. Sci. 2007. Vol. 26. № 5. P. 686-693.
178. Zhu H., Martin T.M., Ye L., Sedykh A., Young D.M., Tropsha A. Quantitative Structure-Activity Relationship Modeling of Rat Acute Toxicity by Oral Exposure // Chem. Res. Toxicol. 2009. Vol. 22. № 12. P. 1913-1921.
179. Devillers J. Prediction of Mammalian Toxicity of Organophosphorus Pesticides from QSTR Modeling // SAR QSAR Environ. Res. 2004. Vol. 15. № 5-6. P. 501-510.
180. Muskal S.M., Jha S.K., Kishore M.P., Tyagi P. A Simple and Readily Integratable Approach to Toxicity Prediction // J. Chem. Inf. Comput. Sci. 2003. Vol. 43. № 5. P. 1673-1678.
181. Toropov A.A., Rasulev B.F., Leszczynski J. QSAR Modeling of Acute Toxicity by Balance of Correlations // Bioorg. Med. Chem. 2008. Vol. 16. № 11. P. 5999-6008.
182. Tremolada P., Finizio A., Villa S., Gaggi C., Vighi M. Quantitative Inter-Specific Chemical Activity Relationships of Pesticides in the Aquatic Environment // Aquat. Toxicol. 2004. Vol. 67. № 1. P. 87-103.
183. Turabekova M.A., Rasulev B.F., Dzhakhangirov F.N., Salikhov S.I. Aconitum and Delphinium Alkaloids "Drug-likeness" Descriptors Related to Toxic Mode of Action // Environ. Toxicol. Phar. 2008. Vol. 25. № 3. P. 310-320.
184. Yan D., Jiang X., Xu S., Wang L., Bian Y., Yu G. Quantitative Structure-Toxicity Relationship Study of Lethal Concentration to Tadpole (*Bufo vulgaris*

- formosus) for Organophosphorous Pesticides // Chemosphere. 2008. Vol. 71. № 10. P. 1809-1815.
185. Yan D., Jiang X., Yu G., Zhao Z., Bian Y., Wang F. Quantitative Structure–Toxicity Relationships of Organophosphorous Pesticides to Fish (*Cyprinus carpio*) // Chemosphere. 2006. Vol. 63. № 5. P. 744-750.
186. Zvinavashe E., Du T., Griff T., Van den Berg H.H.J., Soffers A.E.M.F., Vervoort J., Murk A.J., Rietjens I.M.C.M. Quantitative Structure-Activity Relationship Modeling of the Toxicity of Organothiophosphate Pesticides to *Daphnia magna* and *Cyprinus carpio* // Chemosphere. 2009. Vol. 75. № 11. P. 1531-1538.
187. Padmanabhan J., Parthasarathi R., Subramanian V., Chattaraj P.K. Group Philicity and Electrophilicity as Possible Descriptors for Modeling Ecotoxicity Applied to Chlorophenols // Chem. Res. Toxicol. 2006. Vol. 19. № 3. P. 356-364.
188. Zhao Y.H., Ji G.D., Cronin M.T.D., Dearden J.C. QSAR Study of the Toxicity of Benzoic Acids to *Vibrio fischeri*, *Daphnia magna* and Carp // Sci. Total Environ. 1998. Vol. 216. № 3. P. 205-215.
189. Dom N., Knapen D., Benoot D., Nobels I., Blust R. Aquatic Multi-Species Acute Toxicity of (Chlorinated) Anilines: Experimental versus Predicted Data // Chemosphere. 2010. Vol. 81. № 2. P. 177-186.
190. Lin K.H., Jaw C.G., Yen J.H., Wang Y.S. Molecular Connectivity Indices for Predicting Bioactivities of Substituted Nitrobenzene and Aniline Compounds // Ecotoxicol. Environ. Saf. 2009. Vol. 72. № 7. P. 1942-1949.
191. Bowen K.R., Flanagan K.B., Acree Jr W.E., Abraham M.H. Correlating Toxicities of Organic Compounds to Select Protozoa Using the Abraham Model // Sci. Total Environ. 2006. Vol. 369. № 1-3. P. 109-118.
192. Reuschenbach P., Silvani M., Dammann M., Warnecke D., Knacker T. ECOSAR Model Performance with a Large Test Set of Industrial Chemicals // Chemosphere. 2008. Vol. 71. № 10. P. 1986-1995.

193. Ashauer R., Hintermeister A., Potthoff E., Escher B.I. Acute Toxicity of Organic Chemicals to *Gammarus Pulex* Correlates with Sensitivity of *Daphnia Magna* Across Most Modes of Action // *Aquat. Toxicol.* 2011. Vol. 103. № 1-2. P. 38-45.
194. Devillers J. A General QSAR Model for Predicting the Acute Toxicity of Pesticides to *Lepomis Macrochirus* // *SAR QSAR Environ. Res.* 2001. Vol. 11. № 5. P. 397-417.
195. Cronin M.T.D., Dearden J.C., Duffy J.C., Edwards R., Manga N., Worth A.P., Worgan A.D.P. The Importance of Hydrophobicity and Electrophilicity Descriptors in Mechanistically-Based QSARs for Toxicological Endpoints // *SAR QSAR Environ. Res.* 2002. Vol. 13. № 1. P. 167-176.
196. Furuhashi A., Toida T., Nishikawa N., Aoki Y., Yoshioka Y., Shiraishi H. Development of an Ecotoxicity QSAR Model for the KAsinhou Tool for Ecotoxicity (KATE) System, March 2009 Version // *SAR QSAR Environ. Res.* 2010. Vol. 21. № 5-6. P. 403-413.
197. Hoover K.R., Acree Jr. W.E., Abraham M.H. Chemical Toxicity Correlations for Several Fish Species Based on the Abraham Solvation Parameter Model // *Chem. Res. Toxicol.* 2005. Vol. 18. № 9. P. 1497-1505.
198. Lessigiarska I., Worth A.P., Sokull-Kluttgen B., Jeram S., Dearden J.C., Netzeva T.I., Cronin M.T.D. QSAR Investigation of a Large Data Set for Fish, Algae and *Daphnia* Toxicity // *SAR QSAR Environ. Res.* 2004. Vol. 15. № 5-6. P. 413-431.
199. Huang C.P., Wang Y.J., Chen C.Y. Toxicity and Quantitative Structure–Activity Relationships of Nitriles Based on *Pseudokirchneriella Subcapitata* // *Ecotoxicol. Environ. Saf.* 2007. Vol. 67. № 3. P. 439-446.
200. Huang H., Wang X., Ou W., Zhao J., Shao Y., Wang L. Acute Toxicity of Benzene Derivatives to the Tadpoles (*Rana Japonica*) and QSAR Analyses // *Chemosphere.* 2003. Vol. 53. № 8. P. 963-970.

201. Wang X., Dong Y., Wang L., Han S. Acute Toxicity of Substituted Phenols to *Rana Japonica* Tadpoles and Mechanism-Based Quantitative Structure-Activity Relationship (QSAR) Study // *Chemosphere*. 2001. Vol. 44. № 3. P. 447-455.
202. Roy K., Ghosh G. QSTR with Extended Topochemical Atom (ETA) Indices. VI. Acute Toxicity of Benzene Derivatives to Tadpoles (*Rana japonica*) // *J. Mol. Model*. 2006. Vol. 12. № 3. P. 306-316.
203. Roy K., Sanyal I. QSTR with Extended Topochemical Atom Indices. 7. QSAR of Substituted Benzenes to *Saccharomyces cerevisiae* // *QSAR Comb. Sci*. 2006. Vol. 25. № 4. P. 359-371.
204. Wang X., Yin C., Wang L. Structure–Activity Relationships and Response–Surface Analysis of Nitroaromatics Toxicity to the Yeast (*Saccharomyces Cerevisiae*) // *Chemosphere*. 2002. Vol. 46. № 7. P. 1045-1051.
205. Lu G.H., Yuan X., Zhao Y.H. QSAR Study on the Toxicity of Substituted Benzenes to the Algae (*Scenedesmus Obliquus*) // *Chemosphere*. 2001. Vol. 44. № 3. P. 437-440.
206. Lu G.H., Wang C., Guo X.L. Prediction of Toxicity of Phenols and Anilines to Algae by Quantitative Structure-Activity Relationship // *Biomed. Environ. Sci*. 2008. Vol. 21. № 3. P. 193-196.
207. Wang C., Lu G., Tang Z., Guo X. Quantitative Structure-Activity Relationships for Joint Toxicity of Substituted Phenols and Anilines to *Scenedesmus Obliquus* // *J. Environ. Sci*. 2008. Vol. 20. № 1. P. 115-119.
208. Zhao Y.H., Qin W.C., Su L.M., Yuan X., Lu G.H., Abraham H.M. Toxicity of Substituted Benzenes and Algae (*Scenedesmus Obliquus*) with Solvation Equation // *Chinese Sci. Bull*. 2009. Vol. 54. № 10. P. 1690-1696.
209. Sacan M.T., Ozkul M., Erdem S.S. *Chemosphere*. 2007. Vol. 68. № 4. P. 695-702.
210. Schmitt H., Altenburger R., Jastorff B., Schuurmann G. Quantitative Structure-Activity Analysis of the Algae Toxicity of Nitroaromatic Compounds // *Chem. Res. Toxicol*. 2000. Vol. 13. № 6. P. 441-450.

211. Gramatica P., Vighi M., Consolaro F., Todeschini R., Finizio A., Faust M. QSAR Approach for the Selection of Congeneric Compounds with a Similar Toxicological Mode of Action // *Chemosphere*. 2001. Vol. 42. № 8. P. 873-883.
212. Ren S., Frymier P.D. Estimating the Toxicities of Organic Chemicals to Bioluminescent Bacteria and Activated Sludge // *Water Res.* 2002. Vol. 36. № 17. P. 4406-4414.
213. Ren S., Frymier P.D. Modeling the Toxicity of Polar and Nonpolar Narcotic Compounds to Luminescent Bacterium Shk1 // *Environ. Toxicol. Chem.* 2002. Vol. 21. № 12. P. 2649-2653.
214. Hemmateenejad B., Mehdipour A.R., Miri R., Shamsipur M. Comparative QSAR Studies on Toxicity of Phenol Derivatives Using Quantum Topological Molecular Similarity Indices // *Chem. Biol. Drug Des.* 2010. Vol. 75. № 5. P. 521-531.
215. Lu G.H., Wang C., Li Y.M. QSARS for Acute Toxicity of Halogenated Benzenes to Bacteria in Natural Waters // *Biomed. Environ. Sci.* 2006. Vol. 19. № 6. P. 457-460.
216. Bowen K.R., Flanagan K.B., Acree Jr. W.E., Abraham M.H., Rafols C. Correlation of the Toxicity of Organic Compounds to Tadpoles Using the Abraham Model // *Sci. Total Environ.* 2006 Vol. 371. № 1-3. P. 99-109.
217. Xu J.B., Jing T.S., Pauli W., Berger S. Quantitative Structure-Activity Relationships for the Toxicity of Nitrobenzenes to *Tetrahymena Thermophila* // *J. Environ. Sci. Health, Part A*. 2002. Vol. 37. № 4. P. 563-571.
218. Furusjo E., Svenson A., Rahmberg M., Andersson M. The Importance of Outlier Detection and Training Set Selection for Reliable Environmental QSAR Predictions // *Chemosphere*. 2006. Vol. 63. № 1. P. 99-108.
219. Toropov A.A., Benfenati E. SMILES as an Alternative to the Graph in QSAR Modelling of Bee Toxicity // *Comput. Biol. Chem.* 2007. Vol. 31. № 1. P. 57-60.
220. Boyd E.M., Killham K. Meharg A.A. Toxicity of Mono-, Di- and Tri-Chlorophenols to Lux Marked Terrestrial Bacteria, *Burkholderia* Species Rasc

- c2 and *Pseudomonas Fluorescens* // *Chemosphere*. 2001. Vol. 43. № 2. P. 157-166.
221. Wang X., Yu J., Wang Y., Wang L. Mechanism-Based Quantitative Structure-Activity Relationships for the Inhibition of Substituted Phenols on Germination Rate of *Cucumis Sativus* // *Chemosphere*. 2002. Vol. 46. № 2. P. 241-250.
222. Roy K., Ghosh G. QSTR with Extended Topochemical Atom (ETA) Indices 8. QSAR for the Inhibition of Substituted Phenols on Germination Rate of *Cucumis Sativus* Using Chemometric Tools // *QSAR Comb. Sci.* 2006. Vol. 25. № 10. P. 846-859.
223. Wang X., Sun C., Wang Y., Wang L. Quantitative Structure-Activity Relationships for the Inhibition Toxicity to Root Elongation of *Cucumis Sativus* of Selected Phenols and Interspecies Correlation with *Tetrahymena Pyriformis* // *Chemosphere*. 2002. Vol. 46. № 2. P. 153-161.
224. Lee H.J., Villaume J., Cullen D.C., Kim B.C., Gu M.B. Monitoring and Classification of PAH Toxicity Using an Immobilized Bioluminescent Bacteria // *Biosens. Bioelectron.* 2003. Vol. 18. № 5-6. P. 571-577.
225. Devillers J., Pham-Delegue M.H., Decourtye A., Budzinski H., Cluzeau S., Maurin G. Structure-Toxicity Modeling of Pesticides to Honey Bees // *SAR QSAR Environ. Res.* 2002. Vol. 13. № 7-8. P. 641-648.
226. Grodnitzky J.A., Coats J.R. QSAR Evaluation of Monoterpenoids' Insecticidal Activity // *J. Agric. Food Chem.* 2002. Vol. 50. № 16. P. 4576-4580.
227. Lessigiarska I., Worth A.P., Netzeva T.I., Dearden J.C., Cronin M.T.D. Quantitative Structure–Activity–Activity and Quantitative Structure–Activity Investigations of Human and Rodent Toxicity // *Chemosphere*. 2006. Vol. 65. № 10. P. 1878-1887.
228. Bundy J.G., Morriss A.W.J., Durham D.G., Campbell C.D., Paton G.I. Development of QSARs to Investigate the Bacterial Toxicity and Biotransformation Potential of Aromatic Heterocyclic Compounds // *Chemosphere*. 2001. Vol. 42. № 8. P. 885-892.

229. Piclin N., Pintore M., Wechman C., Roncaglioni A., Benfenati E., Chretien J.R. Ecotoxicity Prediction by Adaptive Fuzzy Partitioning: Comparing Descriptors Computed on 2D and 3D Structures // SAR QSAR Environ. Res. 2006. Vol. 17. № 2. P. 225-251.
230. Lessigiarska I., Cronin M.T.D., Worth A.P., Dearden J.C., Netzeva T.I. QSARs for Toxicity to the Bacterium *Sinorhizobium Meliloti* // SAR QSAR Environ. Res. 2004. Vol. 15. № 3. P. 169-190.
231. Tichý M., Bořek-Dohalský V., Matoušová D., Rucki M., Feltl L., Roth Z. Prediction of Acute Toxicity of Chemicals in Mixtures: Worms *Tubifex tubifex* and Gas/Liquid Distribution // SAR QSAR Environ. Res. 2002. Vol. 13. № 2. P. 261-269.
232. Tichy M., Rucki M., Hanzlikova I., Roth Z. Acute Toxicity Estimation by Calculation – Tubifex Assay and Quantitative Structure-Activity Relationships // Environ. Toxicol. Chem. 2008. Vol. 27. № 11. P. 2281-2286.
233. Tichy M., Pokorna A., Hanzlikova I., Nerudova J., Tumova J., Uzlova R. Primary Rat Hepatocytes in Chemical Testing and QSAR Predictive Applicability // Toxicol. In Vitro. 2010. Vol. 24. № 1. P. 240-244.
234. Cai B., Xie L., Yang D., Arcangeli J.P. Toxicity Evaluation and Prediction of Toxic Chemicals on Activated Sludge System // J. Hazard. Mater. 2010. Vol. 177. № 1-3. P. 414-419.
235. Lee S., Park K., Ahn H.S., Kim D. Importance of Structural Information in Predicting Human Acute Toxicity from in Vitro Cytotoxicity Data // Toxicol. Appl. Pharmacol. 2010. Vol. 246. № 1-2. P. 38-48.
236. Roy K., Ghosh G. QSTR with Extended Topochemical Atom Indices. 10. Modeling of Toxicity of Organic Chemicals to Humans Using Different Chemometric Tools // Chem. Biol. Drug Des. 2008. Vol. 72. № 5. P. 383-394.
237. Hansch C., Kurup A. QSAR of Chemical Polarizability and Nerve Toxicity. 2 // J. Chem. Inf. Comput. Sci. 2003. Vol. 43. № 5. P. 1647-1651.

238. Russom C.L., Bradbury S.P., Broderius S.J., Hammermeister D.E., Drummond R.A. Predicting Modes of Toxic Action from Chemical Structure: Acute Toxicity in the Fathead Minnow (*Pimephales promelas*) // Environ. Toxicol. Chem. 1997. Vol. 16. № 5. P. 948-967.
239. Ellison C.M., Cronin M.T.D., Madden J.C., Schultz T.W. Definition of the Structural Domain of the Baseline Non-polar Narcosis Model for *Tetrahymena pyriformis* // SAR and QSAR in Environ. Res. 2008. Vol. 19. № 7. P. 751-783.
240. De Wolf W., Lieder P.H., Walker J.D. Application of QSARs: Correlation of Acute Toxicity in the Rat Following Oral or Inhalation Exposure // QSAR Comb. Sci. 2004. Vol. 23. № 7. P. 521-525.
241. Freidig A.P., Dekkers S., Verwei M., Zvinavashe E., Bessems J.G.M., van de Sandt J.J.M. Development of a QSAR for Worst Case Estimates of Acute Toxicity of Chemically Reactive Compounds // Toxicol. Lett. 2007. Vol. 170. № 3. P. 214-222.
242. URL: <http://esis.jrc.ec.europa.eu/>.
243. URL: <http://webnet.oecd.org/HPV/UI/Default.aspx>.
244. URL: <http://www.nlm.nih.gov/pubs/factsheets/toxnetfs.html>.
245. URL: <http://apps.kemi.se/nclass/>.
246. URL: <http://apps.kemi.se/riskline/>.
247. URL: <http://accelrys.com/products/databases/bioactivity/toxicity.html>.
248. URL: <http://www.vet.utk.edu/TETRATOX/index.php>.
249. URL: <http://openmopac.net/>.
250. URL: <http://accelrys.com/products/informatics/desktop-software.html>.
251. URL: <http://www.semichem.com/>.
252. URL: <http://www.tripos.com/>.
253. URL: <http://www.oasis-lmc.org/>.
254. URL: <http://www.gaussian.com/>.
255. URL: <http://accelrys.com/>.
256. URL: <http://www.biobyte.com/>.

257. URL: <http://www.epa.gov/oppt/exposure/pubs/episuite.htm>.
258. Argese E., Bettiol C., Fasolo M., Zambon A., Agnoli F. Substituted Aniline Interaction with Submitochondrial Particles and Quantitative Structure-Activity Relationships // *Biochim. Biophys. Acta*. 2002. Vol. 1558. № 2. P. 151-160.
259. Struck S., Schmidt U., Gruening B., Jaeger I.S., Hossbach J., Preissner R. Toxicity vs Potency: Elucidation of Toxicity Properties Discriminating Between Toxins, Drugs, and Natural Compounds // *Genome Inform.* 2008. Vol. 20. P. 231-242.
260. Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings // *Adv. Drug Deliv. Rev.* 2001. Vol. 46. № 1-3. P. 3-26.
261. Niu B., Jin Y., Lu W., Li G. Predicting Toxic Action Mechanisms of Phenols Using AdaBoost Learner // *Chemometr. Intell. Lab. Syst.* 2009. Vol. 96. № 1. P. 43-48.
262. Ren S. Predicting Three Narcosis Mechanisms of Aquatic Toxicity // *Toxicol. Lett.* 2002. Vol. 133. № 1-2. P. 127-139.
263. Ren S. Classifying Class I and Class II Compounds By Hydrophobicity and Hydrogen Bonding Descriptors // *Environ. Toxicol.* 2002. Vol. 17. № 5. P. 415-423.
264. Spycher S., Nendza M., Gasteiger G. Comparison of Different Classification Methods Applied to a Mode of Toxic Action Data Set // *QSAR Comb. Sci.* 2004. Vol. 23. № 9. P. 779-791.
265. Xu L., Wang X., Zhao W. Bridging the Gap Between Molecular Descriptors and Mechanism: Cases Studies by Molecular Dynamics Simulations // *J. Mol. Graph. Model.* 2009. Vol. 27. № 7. P. 829-835.
266. Toropov A.A., Benfenati E. QSAR Modelling of Aldehyde Toxicity by Means of Optimisation of Correlation Weights of Nearest Neighbouring Codes // *J. Mol. Struct. (Theochem)*. 2004. Vol. 676. № 1-3. P. 165-169.

267. Khadikar P.V., Karmarkar S., Singh S., Shrivastava A. Use of the PI Index in Predicting Toxicity of Nitrobenzene Derivatives // *Bioorg. Med. Chem.* 2002. Vol. 10. № 10. P. 3163-3170.
268. Verhaar H.J.M., Van Leeuwen C.J., Hermens J.L.M. Classifying Environmental Pollutants // *Chemosphere*. 1992. Vol. 25. № 4. P. 471-491.
269. Bradbury S.P. Predicting Modes of Toxic Action from Chemical Structure: An Overview // *SAR QSAR Environ. Res.* 1994. Vol. 2. № 1-2. P. 89-104.
270. Broderius S., Kahl M. Acute Toxicity of Organic Chemical Mixtures to the Fathead Minnow // *Aquat. Toxicol.* 1985. Vol. 6. № 4. P. 307-322.
271. Schultz T.W., Holcombe G.W., Phipps G.L. Relationships of Quantitative Structure–Activity of Selected Phenols in the *Pimephales promelas* and *Tetrahymena pyriformis* Test Systems // *Ecotoxicol. Environ. Saf.* 1986. Vol. 12. № 2. P. 146-153.
272. Jaworska J.S., Schultz T.W. Quantitative Relationships of Structure-Activity and Volume Fraction for Selected Nonpolar and Polar Narcotic Chemicals // *SAR QSAR Environ. Res.* 1993. Vol. 1. № 1. P. 3-19.
273. Basak S.C., Grunwald G.D., Host G.E., Niemi G.J., Bradbury S.P. A Comparative Study of Molecular Similarity, Statistical, and Neural Methods for Predicting Toxic Modes of Action // *Environ. Toxicol. Chem.* 1998. Vol. 17. № 6. P. 1056-1064.
274. McKim J.M., Bradbury S.P., Niemi G.J. Fish Acute Toxicity Syndromes and Their Use in the QSAR Approach to Hazard Assessment // *Environ. Health Perspect.* 1987. Vol. 71. P. 171-186.
275. Nendza M., Müller M. Discriminating Toxicant Classes by Mode of Action: 2. Physico-Chemical Descriptors // *Quant. Struct.-Act. Relat.* 2000. Vol. 19. № 6. P. 581-598.
276. Nendza M., Wenzel A. Discriminating Toxicant Classes by Mode of Action 1. (Eco)toxicity Profiles // *Environ. Sci. Pollut. Res.* 2006. Vol. 13. № 3. P. 192-203.

277. Sanderson H., Thomsen M. Ecotoxicological Quantitative Structure–Activity Relationships for Pharmaceuticals // *Bull. Environ. Contam. Toxicol.* 2007. Vol. 79. № 3. P. 331-335.
278. Raimondo S., Mineau P., Barron M.G. Estimation of Chemical Toxicity to Wildlife Species Using Interspecies Correlation Models // *Environ. Sci. Technol.* 2007. Vol. 41. № 16. P. 5888-5894.
279. Raimondo S., Jackson C.R., Barron M.G. Influence of Taxonomic Relatedness and Chemical Mode of Action in Acute Interspecies Estimation Models for Aquatic Species // *Environ. Sci. Technol.* 2010. Vol. 44. № 19. P. 7711-7716.
280. Rubach M.N., Baird D.J., Van den Brink P.J. A New Method for Ranking Mode-Specific Sensitivity of Freshwater Arthropods to Insecticides and Its Relationship to Biological Traits // *Environ. Toxicol. Chem.* 2010. Vol. 29. № 2. P. 476-487.
281. Ramos E.U., Vaes W.H.J., Verhaar H.J.M., Hermens J.L.M. Quantitative Structure–Activity Relationships for the Aquatic Toxicity of Polar and Nonpolar Narcotic Pollutants // *J. Chem. Inf. Comput. Sci.* 1998. Vol. 38. № 5. P. 845-852.
282. Ramos E.U., Vaes W.H.J., Mayer P., Hermens J.L.M. Algal Growth Inhibition of *Chlorella pyrenoidosa* by Polar Narcotic Pollutants: Toxic Cell Concentrations and QSAR Modeling // *Aquat. Toxicol.* 1999. Vol. 46. № 1. P. 1-10.
283. Roy P.P., Paul S., Mitra I., Roy K. On Two Novel Parameters for Validation of Predictive QSAR Models // *Molecules.* 2009. Vol. 14. № 5. P. 1660-1701.
284. Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation // *Mol. Inf.* 2010. Vol. 29. № 6-7. P. 476-488.
285. Gramatica P. Principles of QSAR Models Validation: Internal and External // *QSAR Comb. Sci.* 2007. Vol. 26. № 5. P. 694-701.
286. URL: <http://www.oecd.org/dataoecd/33/37/37849783.pdf/>.

287. Eriksson L., Jaworska J., Worth A.P., Cronin M.T.D., McDowell R.M., Gramatica P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs // Environ. Health Perspect. 2003. Vol. 111. № 10. P. 1361-1375.
288. Netzeva T.I., Worth A.P., Aldenberg T., Benigni R., Cronin M.T.D., Gramatica P., Jaworska J.S., Kahn S., Klopman G., Marchant C.A., Myatt G., Nikolova-Jeliazkova N., Patlewicz G.Y., Perkins R., Roberts D.W., Schultz T.W., Stanton D.T., Van de Sandt J.J.M., Tong W., Veith G., Yang C. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure–Activity Relationships // ATLA. 2005. Vol. 33. № 2. P. 1-19.
289. Könemann H. Quantitative Structure-Activity Relationships in Fish Toxicity Studies Part 1: Relationship for 50 Industrial Pollutants // Toxicol. 1981. Vol. 19. № 3. P. 209-221.
290. Veith G.D., Call D.J., Brooke L.T. Structure - Toxicity Relationships for the Fathead Minnow, *Pimephales promelas*: Narcotic Industrial Chemicals // Can. J. Fish. Aquat. Sci. 1983. Vol. 40. № 6. P. 743-748.
291. Roberts D.W., Costello J.F. Mechanisms of Action for General and Polar Narcosis: A Difference in Dimension // QSAR Comb. Sci. 2003. Vol. 22. № 2. P. 226-233.
292. Vaes W.H.J., Ramos E.U., Verhaar H.J.M., Hermens J.L.M. Acute Toxicity of Nonpolar Versus Polar Narcosis: Is There a Difference? // Environ. Toxicol. Chem. 1998. Vol. 17. № 7. P. 1380-1384.
293. Mekenyan O.G., Veith G.D. Relationships Between Descriptors for Hydrophobicity and Soft Electrophilicity in Predicting Toxicity // SAR QSAR Environ. Res. 1993. Vol. 1. № 4. P. 335-344.
294. Cronin M.T.D., Schultz T.W. Structure-Toxicity Relationships for Phenols to *Tetrahymena pyriformis* // Chemosphere. 1996. Vol. 32. № 8. P. 1453-1468.
295. Sanderson H., Johnson D.J., Wilson C.J., Brain R.A., Solomon K.R. Probabilistic Hazard Assessment of Environmentally Occurring Pharmaceuticals

- Toxicity to Fish, Daphnids and Algae by ECOSAR Screening // *Toxicol. Lett.* 2003. Vol. 144. № 3. P. 383-395.
296. Sanderson H., Johnson D.J., Reitsma T., Brain R.A., Wilson C.J., Solomon K.R. Ranking and Prioritization of Environmental Risks of Pharmaceuticals in Surface Waters // *Regul. Toxicol. Pharmacol.* 2004. Vol. 39. № 2. P. 158-183.
297. Lagunin A., Zakharov A., Filimonov D., Poroikov V. QSAR Modelling of Rat Acute Toxicity on the Basis of PASS Prediction // *Mol. Inf.* 2011. Vol. 30. № 2-3. P. 241-250.
298. URL: <http://accelrys.com/>.
299. URL: <http://www.epa.gov/>.
300. URL: <http://www.multicase.com/>.
301. URL: <http://www.epa.gov/>.
302. URL: <http://www.oasis-lmc.org/>.
303. URL: <http://www.pharmaexpert.ru/passonline/>.
304. URL: <http://kate.nies.go.jp/>.
305. Раевский О.А., Григорьев В.Ю., Трепалин С.В. Свидетельство об официальной регистрации программы для ЭВМ HYBOT (Hydrogen Bond Thermodynamics) № 990090 от 26 февраля 1999 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.
306. Hansch C., Leo A., Hoekman D. Exploring QSAR: Hydrophobic, Electronic and Steric Constants. Washington: ACS, 1995. 368 p.
307. Katritzky A.R., Oliferenko A.A., Oliferenko P.V., Petrukhin R., Tatham D.B., Maran U., Lomaka A., Acree W.E. A General Treatment of Solubility. 1. The QSPR Correlation of Solvation Free Energies of Single Solutes in Series of Solvents // *J. Chem. Inf. Comput. Sci.* 2003. Vol. 43. № 6. P. 1794–1805.
308. Свойства органических соединений. Справочник / Под ред. А.А. Потехина. Л.: Химия, 1984. 520 с.
309. Рид Р., Праусниц Дж., Шервуд Т. Свойства газов и жидкостей. Л.: Химия, 1982. 592 с.

310. Cronin M.T.D., Schultz T.W. Validation of *Vibrio fischeri* Acute Toxicity Data: Mechanism of Action-based QSARs for Non-polar Narcotics and Polar Narcotic Phenols // The Science of the Total Environment. 1997. Vol. 204. № 1. P. 75-88.
311. Zhu H., Tropsha A., Fourches D., Varnek A., Papa E., Gramatica P., Öberg O., Dao P., Cherkasov A., Tetko I. Combinatorial QSAR Modeling of Chemical Toxicants Tested against *Tetrahymena pyriformis* // J. Chem. Inf. Model. 2008. Vol. 48. № 4. P. 766-784.
312. URL: <http://cfpub.epa.gov/ecotox/>.
313. URL: <http://accelrys.com/products/databases/bioactivity/rtecs.html>.
314. Муромцев Г.С., Кокурин А.В. Методы определения регуляторов роста растений (методические рекомендации). М.: ВАСХНИЛ, 1984. 21 с.
315. URL: <http://www.taletе.mi.it/>.
316. Raevsky O., Sapegin A., Zefirov N. QSAR Discriminant-Regression Model // In: QSAR: Rational Approaches to the Design of Bioactive Compounds / Silipo C., Vittoria A., Eds. Amsterdam: Elsevier, 1991. P. 189-192.
317. Raevsky O.A., Trepalina E.P., Trepalin S.V. SLIPPER - a New Program for Water Solubility, Lipophilicity and Permeability Prediction // In: Molecular Modeling and Prediction of Bioactivity / Gundertofe K., Jorgensen F., Eds. New York, Boston, Dordrecht: Kluwer/Plenum Publishers, 2000. P. 489-490.
318. Livingstone D.J. Data Analysis for Chemists: Applications to QSAR and Chemical Product Design. Oxford: Oxford University Press, 1995. P. 48-64.
319. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. С. 153-155.
320. URL: <http://www.knowitall.com/>.
321. URL: <http://www.hyper.com/>.
322. Новиков В.П., Раевский О.А. Представление молекулярной структуры в виде спектра межатомных расстояний для изучения связи структура - биологическая активность // Хим.-фарм. журн. 1982. Т. 16. № 5. С. 574-581.

323. Gerasimenko V.A., Trepalin S.V., Raevsky O.A. MOLDIVS - A New Program for Molecular Similarity and Diversity Calculations // In: Molecular Modeling and Prediction of Bioactivity / Gundertofte K., Jorgensen F.S., Eds. New York: Kluwer Academic/Plenum Publishers, 2000. P. 423–424.
324. Murtagh F., Heck A. Multivariate Data Analysis, Astrophysics and Space Science Library. Dordrecht: D. Reidel Publ. Co., 1987. 210 p.
325. URL: <http://www.spss.ru/>.
326. URL: <http://software.intel.com/en-us/articles/visual-numerics-imsl-fortran-library/>.
327. Raevsky O.A., Dearden J.C. Creation of Predictive Models of Aquatic Toxicity of Environmental Pollutants with Different Mechanisms of Action on the Basis of Molecular Similarity and HYBOT Descriptors // SAR QSAR Environ. Res. 2004. Vol. 15. № 5-6. P. 433-448.
328. URL: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
329. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. М.: Мир, 1980. 280 с.
330. Gardner M.J., Altman D.J. Calculating Confidence Intervals for Proportions and Their Differences // In: Statistics with Confidence / Gardner M.J., Altman D.G., Eds. London: BMJ Publishing Group, 1989. P. 28-33.
331. Matthews B.W. Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme // Biochim. Biophys. Acta. 1975. Vol. 405. № 2. P. 442-451.
332. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. 456 с.
333. Wold S., Eriksson L. Validation Tools // In: Chemometric Methods in Molecular Design / van de Waterbeemd H., Ed. Weinheim: VCH, 1995. P. 309-318.
334. URL: <http://ambit.sourceforge.net/>.
335. Tropsha A., Gramatica P., Gombar V.K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and

Interpretation of QSPR Models // QSAR Comb. Sci. 2003. Vol. 22. № 1. P. 69-77.

336. Deranleau D.A. Theory of the Measurement of Weak Molecular Complexes. 1. General Considerations // J. Am. Chem. Soc. 1969. Vol. 91. № 15. P. 4044-4049.
337. Новиков В.П., Игнатьева Т.И., Раевский О.А. Обработка потенциометрических данных по универсальной схеме расчета равновесий в растворах на примере анализа комплексообразования метилфосфоновой кислоты // Журн. неорг. химии. 1986. Т. 31. № 6. С. 1474-1478.
338. Химмельблау Д.М. Прикладное нелинейное программирование. М.: Мир, 1975. С. 448.
339. Хартли Ф., Бергес К., Оллок Р. Равновесия в растворах. М.: Мир, 1983. 360 с.
340. Гордон А., Форд Р. Спутник химика. М.: Мир, 1976. 541 с.
341. Соколов В.Б., Иванов А.Н., Епишина Т.А., Мартынов И.В. Взаимодействие О-алкилметилфосфонитов с 1,1-дихлорнитропропаном // Журн. общ. химии. 1987. Т. 57. № 4. С. 952-953.
342. Григорьев В.Ю., Соловьев В.П., Додонов М.В., Брель В.К., Мартынов И.В., Раевский О.А. Электронодонорные и акцепторные функции физиологически активных и модельных соединений. XIV. Электроноакцепторная способность группы N-H дифторнитроацетанилидов // Журн. общ. химии. 1989. Т. 59. № 4. С. 846-851.
343. Полинг Л., Полинг П. Химия. М.: Мир, 1978. 683 с.
344. Hadzi D. // In: Spectroscopy of Biological Molecules / Sandorfy C., Teofanidis C., Eds. New York: Reidel. Publ. Co., 1984. P. 61-85.
345. Naray-Szabo G., Surjan P. // In: Theoretical Chemistry of Biological Systems / Naray-Szabo G., Ed. Amsterdam: Elsevier, 1986. P. 1-100.

346. Hayward R.C. Abiotic receptors // Chem. Soc. Rev. 1983. Vol. 12. № 3. P. 285-308.
347. Guidry R.M., Drago R.S. An Extension of the E and C equation to Evaluate Constant Contributions to a Series of Observed "Enthalpies of Adduct Formation" // J. Am. Chem. Soc. 1973. Vol. 95. № 3. P. 759-763.
348. Drago R.S., Parr L.B., Chamberlain C.S. Solvent Effects and Their Relationship to the E and C equation // J. Am. Chem. Soc. 1977. Vol. 99. № 10. P. 3203-3209.
349. Kroeger M.K., Drago R.S. Quantitative Prediction and Analysis of Enthalpies for the Interaction of Gas-phase Ion-ion, Gas-phase Ion-molecule, and Molecule-molecule Lewis Acid-base Systems // J. Am. Chem. Soc. 1981. Vol. 103. № 12. P. 3250-3262.
350. Иогансен А.В. Зависимость между энергией водородной связи и интенсивностью инфракрасного поглощения // Докл. АН СССР. 1965. Т. 164. № 3. С. 610-613.
351. Иогансен А.В., Куркчи Г.А., Левина О.В. Газо-хроматографическое исследование галогенводородов // Журн. физ. химии. 1969. Т. 43. № 11. С. 2915-2920.
352. Иогансен А.В. Правило произведения кислотно-основных функций молекул при их ассоциации водородными связями в растворах в CCl_4 // Теор. exper. химия. 1971. Т. 7. № 3. С. 302-311.
353. Иогансен А.В. Оценка влияния среды на свойства водородных связей по правилу произведения кислотно-основных функций молекул // Теор. exper. химия. 1971. Т. 7. № 3. С. 312-317.
354. Терентьев В.А. Термодинамика водородной связи. Саратов: Изд.-во Саратов. ун-та, 1973. 258 с.
355. Раевский О.А., Новиков В.П. Унификация характеристик донорно-акцепторного взаимодействия в рамках проблемы связи структуры с активностью // Хим.-фарм. журн. 1982. Т. 16. № 5. С. 583-586.

356. Раевский О.А., Авидон В.В., Новиков В.П. Использование унифицированной шкалы донорно-акцепторных взаимодействий для анализа сходства структур биологически активных соединений // Хим.-фарм. журн. 1982. Т. 16. № 8. С. 968-971.
357. Bordzilovskii V.Ya., Platonov A.P. Thermochemistry of Methyl *N*-Phenylcarbamate Solutions Containing Aliphatic Alcohols // Russ. J. Gen. Chem. 1996. Vol. 66. № 10. P. 1573-1576.
358. Abraham M.H., Grellier P.L., Prior D.V., Duce P.P., Morris J.J., Taylor P.J. Hydrogen Bonding. Part 7. A Scale of Solute Hydrogen-Bond Acidity Based on log *K* Values for Complexation in Tetrachloromethane // J. Chem. Soc. Perkin Trans. 2. 1989. № 6. P. 699-711.
359. Maria P.C., Gal J.F., Franceschi J.D., Fargin E. Chemometrics of Solvent Basicity: Multivariate Analysis of the Basicity Scales Relevant to Nonprotogenic Solvents // J. Amer. Chem. Soc. 1987. Vol. 109. № 2. P. 483-492.
360. Abraham M.H., Grellier P.L., Prior D.V., Taft R.W., Morris J.J., Taylor P.J., Laurence C., Berthelot M., Doherty R.M., Kamlet M.G., Abboud J.L., Sraidi K., Gutheneuf G. A General Treatment of Hydrogen Bond Complexation Constants in Tetrachloromethane // J. Amer. Chem. Soc. 1988. Vol. 110. № 25. P. 8534-8536.
361. Раевский О.А., Григорьев В.Ю., Соловьев В.П. Моделирование связи структура – активность. II. Оценка электронодонорных и акцепторных функций активных центров в молекулах физиологически активных веществ // Хим.-фарм. журн. 1989. Т. 23. № 11. С. 1294-1300.
362. Abraham M.H., Duce P.P., Prior D.V., Barratt D.G., Morris J.J., Taylor P.J. Hydrogen Bonding. Part 9. Solute Proton Donor and Proton Acceptor Scales for Use in Drug Design // J. Chem. Soc. Perkin Trans. 2. 1989. № 10. P. 1355-1375.
363. Abraham M.H., Grellier P.L., Prior D.V., Morris J.J., Taylor P.J. Hydrogen Bonding. Part 10. A Scale of Solute Hydrogen-Bond Basicity Using log *K*

Values for Complexation in Tetrachloromethane // J. Chem. Soc. Perkin Trans. 2. 1990. № 4. P. 521-529.

364. Мельников Н.Н., Новожилов К.В., Белан С.Р., Пылова Т.Н. Справочник по пестицидам. М.: Химия, 1985. 351 с.
365. Lipophilicity in Drug Action and Toxicology / Pliska V., Testa B., Waterbeemd H., Eds. Weinheim: VCH, 1996. 438 p.
366. Rekker R.F. The Hydrophobic Fragmental Constant: Its Derivation and Application with a Means of Characterizing Membrane Systems. Amsterdam: Elsevier, 1977. 390 p.
367. Rekker R.F., Mannhold R. Calculation of Drug Lipophilicity. Weinheim: VCH, 1992. 112 p.
368. Klopman G., Li J.-Y., Wang S., Dimayuga M. Computer Automated log P Calculations Based on an Extended Group Contribution Approach // J. Chem. Inf. Comput. Sci. 1994. Vol. 34. № 4. P. 752-781.
369. Ghose A.K., Crippen G.M. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure-Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions // J. Chem. Inf. Comput. Sci. 1987. Vol. 27. № 1. P. 21-35.
370. Suzuki T., Kudo Y. Automatic log P Estimation Based on Combined Additive Modeling Methods // J. Comp. Aided Mol. Design. 1990. Vol. 4. № 2. P. 155-198.
371. Niemi G.J., Basak S.C., Grunwald G., Veith G.D. Prediction of Octanol/Water Partition Coefficient (K_{OW}) with Algorithmically Derived Variables // Environ. Toxicol. Chem. 1992. Vol. 11. № 7. P. 893-900.
372. Moriguchi I., Hirono S., Liu Q., Nakagome I., Matsushita Y. Simple Method of Calculating Octanol/Water Partition Coefficient // Chem. Pharm. Bull. 1992. Vol. 40. № 1. P. 127-130.

373. Leahy D.E. Intrinsic Molecular Volume as a Measure of the Cavity Term in Linear Solvation Energy Relationships: Octanol-Water Partition Coefficients and Aqueous Solubilities // *J. Pharm. Sci.* 1986. Vol. 75. № 7. P. 629-636.
374. Waterbeemd H., Mannhold R. Programs and Methods for Calculation of log P-values // *Quant. Struct.-Act. Relat.* 1996. Vol. 15. № 5. P. 410-412.
375. Waterbeemd H., Testa B. The Parametrization of Lipophilicity and Other Structural Properties in Drug Design // *Adv. Drug Res.* 1987. Vol. 16. P. 85-225.
376. Tayar N.E., Testa B. Polar Intermolecular Interactions Encoded in Partition Coefficients // In: *Trends in QSAR and Molecular Modelling 92* / Wermuth C.G., Ed. Leiden: ESCOM, 1993. P. 101-108.
377. Seyler P. Interconversion of Lipophilicities from Hydrocarbon/Water Systems into the Octanol/Water System // *Eur. J. Med. Chem. Chim. Ther.* 1974. Vol. 9. P. 473-479.
378. Moriguchi I., Kanada Y., Komatsu K. Van der Waals Volume and the Related Parameters for Hydrophobicity in Structure-Activity Studies // *Chem. Pharm Bull.* 1976. Vol. 24. № 8. P. 1799-1806.
379. Trepalin S.V., Yarkov A.V., Dolmatova L.M., Zefirov N.S., Finch S.A.E. WinDat: An NMR Database Compilation Tool, User Interface, and Spectrum Libraries for Personal Computers // *J. Chem. Inf. Comput. Sci.* 1995. Vol. 35. № 3. P. 405-411.
380. Lipnick R.L. Narcosis: Fundamental and Baseline Toxicity Mechanism for Nonelectrolyte Organic Chemicals // In: *Practical Application of Quantitative Structure – Activity Relationships (QSAR) in Environmental Chemistry and Toxicology* / Karcher W., Devillers J., Eds. Dordrecht: Kluwer Academic Publishers, 1990. P. 129-144.
381. Raevsky O.A., Perlovich G.L., Schaper K.-J. Physicochemical Properties/Descriptors Governing the Solubility and Partitioning of Chemicals in Water-Solvent-Gas Systems. Part 2. Solubility in 1-octanol // *SAR QSAR Environ. Res.* 2007. Vol. 18. № 5-6. P. 543-578.

382. Vermue M., Sikkema J., Verheul A., Bakker R., Tramper J. Toxicity of Homologous Series of Organic Solvents for the Gram-Positive Bacteria *Arthrobacter* and *Nocardia* Sp. and the Gram-Negative Bacteria *Acinetobacter* and *Pseudomonas* Sp. // Biotechnol. Bioeng. 1993. Vol. 42. № 6. P. 747-758.
383. Hau K.M., Connell D.W., Richardson B.J. Mechanism of Acute Inhalation Toxicity of Alkanes and Aliphatic Alcohols // Environ. Toxicol. Pharmacol. 1999. Vol. 7. № 3. P. 159-167.
384. Альберт А. Избирательная токсичность, Т. 1. М.: Медицина, 1989. С. 58-60.
385. Зенкевич И.Г. Единый закон вариаций любых свойств органических соединений в гомологических рядах // Успехи современного естествознания. 2006. № 7. С. 42-46.
386. Vaal M., Van der Wal J.T., Hermens J., Hoekstra J. Pattern Analysis of the Variation in the Sensitivity of Aquatic Species to Toxicants // Chemosphere. 1997. Vol. 35. № 6. P. 1291-1309.
387. Lipnick R.L. Outliers: Their Origin and Use in the Classification of Molecular Mechanisms of Toxicity // Sci. Total Environ. 1991. Vol. 109/110. P. 131-153.
388. Maggiora G.M. On Outliers and Activity Cliffs – Why QSAR Often Disappoints // J. Chem. Inf. Model. 2006. Vol. 46. № 4. P. 1535.
389. Dunn W.J., Wold S. SIMCA Pattern Recognition and Classification // In: Chemometric Methods in Molecular Design / van de Waterbeemd H., Ed. Weinheim: VCH, 1995. P. 179-193.
390. Раевский О.А., Чистяков В.В., Агабекян Р.С., Сапегин А.М., Зефилов Н.С. Формирование моделей взаимосвязи между структурой фосфорорганических соединений и их способностью к ингибированию ацетилхолинэстераз // Биоорг. химия. 1990. Т. 16. № 11. С. 1509-1522.
391. Guha R., Dutta D., Jurs P.C., Chen T. Local Lazy Regression: Making Use of the Neighborhood to Improve QSAR Predictions // J. Chem. Inf. Model. 2006. Vol. 46. № 4. P. 1836-1847.

392. Gunturi S.B., Archana K., Khandelwal A., Narayanan R. Prediction of hERG Potassium Channel Blockade Using kNN-QSAR and Local Lazy Regression Methods // QSAR Comb. Sci. 2008. Vol. 27. № 11-12. P. 1305-1317.
393. Raevsky O.A. Molecular Lipophilicity Calculations of Chemically Heterogeneous Chemicals and Drugs on the Basis of Structural Similarity and Physicochemical Parameters // SAR QSAR Environ. Res. 2001. Vol. 12. № 4. P. 367-381.
394. Мандельброт Б. Фрактальная геометрия природы. М.: Институт компьютерных исследований, 2002. 656 с.
395. Федер Е. Фракталы. М.: Мир, 1991. 254 с.
396. Risovic D., Mahovic Poljacek S., Gojo M. On Correlation Between Fractal Dimension and Profilometric Parameters in Characterization of Surface Topographies // Appl. Surf. Sci. 2009. Vol. 255. № 7. P. 4283-4288.
397. Eftekhari A. Inspecting Plastic Deformation of Pd by Means of Fractal Geometry // Phys. B: Condens. Matter. 2007. Vol. 387. № 1-2. P. 92-97.
398. Ghanbari Kh., Mousavi M.F., Shamsipur M., Rahmanifar M.S., Heli H. Change in Morphology of Polyaniline/Graphite Composite: A Fractal Dimension Approach // Synth. Met. 2006. Vol. 156. № 14-15. P. 911-916.
399. Koňas P., Buchar J., Severa L. Study of Correlation Between the Fractal Dimension of Wood Anatomy Structure and Impact Energy // Eur. J. Mech. A-Solid. 2009. Vol. 28. № 3. P. 545-550.
400. Lee C.-Y. Mass Fractal Dimension of the Ribosome and Implication of its Dynamic Characteristics // Phys. Rev. E. 2006. Vol. 73. № 4. P. 042901.
401. Mihranyan A., Strømme M. Solubility of Fractal Nanoparticles // Surf. Sci. 2007. Vol. 601. № 2. P. 315-319.
402. Ono Y., Mayama H., Furó I., Sagidullin A.I., Matsushima K., Ura H., Uchiyama T., Tsujii K. Characterization and Structural Investigation of Fractal Porous-Silica over an Extremely Wide Scale Range of Pore Size // J. Coll. Interface Sci. 2009. Vol. 336. № 1. P. 215-225.

403. Artemenko A.G., Kovdienko N.A., Kuz'min V.E., Kamalov G.L., Lozitskaya R.N., Fedchuk A.S., Lozitsky V.P., Dyachenko N.S., Nosach L.N. The Analysis of "Structure – Anticancer Activity" Relationship in a Set of Macrocyclic Pyridinophanes and Their Acyclic Analogues on the Basis of Lattice Model of Molecule Using Fractal Parameters // *Exp. Oncol.* 2002. Vol. 24. P. 123-127.
404. Кроновер Р.М. Фракталы и хаос в динамических системах. Основы теории. М.: Постмаркет, 2000. С. 137.
405. Блатов В.А., Шевченко А.П., Пересыпкина Е.В. Полуэмпирические расчетные методы квантовой химии: Учебное пособие. Самара: Универс-групп, 2005. 32 с.
406. Dimov N., Osman A., Mekenyan O., Papazova D. Selection of Molecular Descriptors Used in Quantitative Structure-Gas Chromatographic Retention Relationships: I. Application to Alkylbenzenes and Naphthalenes // *Analyt. Chim. Acta.* 1994. Vol. 298. № 3. P. 303-317.
407. URL: <http://www.molecularDescriptors.eu/>.
408. Nantasenamat C., Isarankura-Na-Ayudhya C., Naenna T., Prachayasittikul V. A Practical Overview of Quantitative Structure-Activity Relationship // *EXCLI Journal.* 2009. Vol. 8. P. 74-88.
409. Liu P., Long W. Current Mathematical Methods Used in QSAR/QSPR Studies // *Int. J. Mol. Sci.* 2009. Vol. 10. № 5. P. 1978-1998.
410. Razdol'skii A.N., Trepalin S.V., Raevskii O.A. QSAR Modeling Based on Interatomic Interaction Spectra // *Pharm. Chem. J.* 2000. Vol. 34. № 12. С. 654-657.
411. Vračko M., Gasteiger J. A QSAR Study on a Set of 105 Flavonoid Derivatives Using Descriptors Derived From 3D Structures // *Internet Electron. J. Mol. Des.* 2002. Vol. 1. № 10. P. 527-544.
412. Gonzalez M.P., Teran C., Teixeira M., Helguera A.M. QSAR Studies Using Radial Distribution Function for Predicting A1 Adenosine Receptors Agonists // *Bull. Math. Biol.* 2007. Vol. 69. № 1. P. 347-359.

413. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М.: Мир, 1989. 540 с.
414. Preuss R. Very Fast Computation of the Radix-2 Discrete Fourier Transform // IEEE Trans. Acoust. Speech Signal Process. 1982. Vol. 30. № 4. P. 595-607.
415. Miyao T., Arakawa M., Funatsu K. Exhaustive Structure Generation for Inverse-QSPR/QSAR // Mol. Inf. 2010. Vol. 29. № 1-2. P. 111-125.
416. Баскин И.И., Гордеева Е.В., Девдариани Р.О., Зефиоров Н.С., Палюлин В.А., Станкевич М.И. Методология решения обратной задачи в проблеме связи «структура-свойство» для случая топологических индексов // Докл. АН СССР. 1989. Т. 307. № 3. С. 613-616.
417. Mel'nikov A.A., Palyulin V.A., Radchenko E.V., Zefirov N.S. Generation of Chemical Structures on the Basis of QSAR Models of Molecular Field Topology Analysis // Dokl. Chem. 2007. Vol. 415. № 2. P. 196-199.
418. Fink T., Reymond J.-L. Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery // J. Chem. Inf. Model. 2007. Vol. 47. № 2. P. 342-353.
419. Varnek A., Fourches D., Solov'ev V.P., Baulin V.E., Turanov A.N., Karandashev V.K., Fara D., Katritzky A.R. In Silico" Design of New Uranyl Extractants Based on Phosphoryl-Containing Podands: QSPR Studies, Generation and Screening of Virtual Combinatorial Library, and Experimental Tests // J. Chem. Inf. Comput. Sci. 2004. Vol. 44. № 4. P. 1365-1382.
420. Brown N., McKay B., Gasteiger J. The De Novo Design of Median Molecules within a Property Range of Interest // J. Comput. Aided Mol. Des. 2004. Vol. 18. № 12. P. 761-771.
421. Weis D.C., Faulon J.-L., LeBorne R.C., Visco D.P. The Signature Molecular Descriptor. 5. The Design of Hydrofluoroether Foam Blowing Agents Using Inverse-QSAR // Ind. Eng. Chem. Res. 2005. Vol. 44. № 23. P. 8883-8891.

422. Никкел Л.Дж. Регуляторы роста растений. Применение в сельском хозяйстве. М.: Колос, 1984. 192 с.
423. Галкин В.И., Саяхов Р.Д., Черкасов Р.А. Моделирование пространственных эффектов в анализе реакционной способности органических производных непереходных элементов // Металлоорг. химия. 1990. Т.3. С. 986-993.
424. Бредбери М. Концепция гематоэнцефалического барьера. М.: Медицина, 1983. 480 с.
425. Abraham M.H., Ibrahim A., Zhao Y., Acree W.E. A Data Base for Partition of Volatile Organic Compounds and Drugs From Blood/Plasma/Serum to Brain, and an LFER Analysis of the Data // J. Pharm. Sci. 2006. Vol. 95. № 10. P. 2091–2100.
426. Kelder J., Grootenhuis P.D.J., Bayada D.M., Delbressine L.P.C., Ploemen J.-P. Polar Molecular Surface As a Dominating Determinant for Oral Absorption and Brain Penetration of Drugs // Pharm. Res. 1999. Vol. 16. № 10. P. 1514-1519.
427. Mente S.R., Lombardo F. A Recursive-Partitioning Model for Blood-Brain Barrier Permeation // J. Comput. Aided Mol. Des. 2005. Vol. 19. № 7. P. 465-481.